Virtual Screening in Discovery of Novel Kinase Inhibitors

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von
Hongtao Zhao
aus China

Promotionskomitee
Prof. Dr. Amedeo Caflisch (Vorsitz und Leitung der Dissertation)
Dr. Danzhi Huang
Prof. Dr. Cristina Nevado
Prof. Dr. Raimound Dutzler

Zürich 2013

Protein kinases are among the most intensively pursued class of drug targets with inhibitors of about 30 distinct kinase targets in a Phase I clinical trail. The vast majority of these inhibitors are under investigation for the treatment of cancer; however, aberrant kinase function has been implicated in further indications, including immunological, neurological, metabolic and infectious diseases. To date, 14 kinase inhibitors have been approved by the US Food and Drug Administration for the treatment of cancer.

This PhD thesis focuses on the identification of novel tyrosine kinase inhibitors by computational modeling. Over four years, two programs have been developed to complete the virtual screening tool box: Mr. Dock, a fast and accurate flexible-docking tool, and an accurate scoring function. A molecular dynamics-based method (MD-IF) was devised to circumvent issues relating to induced-fit in docking, thus allowing the protein target to be explored by a larger area of chemical space. We further optimized the *in silico* fragment-based approach by introducing a new fragmentation tool called LIBO. The developed methods have been retrospectively validated, and prospectively lead to the discovery of three novel classes of tyrosine kinase inhibitors with activities ranging from low micromolar to nanomolar.

The first pyrrolo[2,3-b]quinoxaline scaffold identified *in silico* formed the basis of a series of derivatives as tyrosine kinase inhibitors, exhibiting potent enzymatic and cellular activity, good solubility, and selectivity profiles comparable to marketed drugs. A second isoquinoline scaffold gave rise to a 160 nM lead compound, which is promising for future optimization. Finally, predicted binding modes were confirmed by X-ray crystallography, indicating that these findings can be ascribed to computational modeling and not serendipity.

Mit ungefär 30 unterschiedlichen Mitgliedern, welche sich aktuell in Phase I – klinischen Studien befinden, gehören die Proteinkinasen zu einer Klasse der meist untersuchten Wirkstoffziele. Die große Mehrheit dieser Kinasen werden in Bezug auf ihre Rolle in Krebserkrankungen untersucht. Dysregulation der Kinasefunktion wird zusätzlich als Faktor in immunologischen, neurologischen, metabolischen und infektiösen Krankheiten angenommen. Bis heute wurden 14 Kinaseinhibitoren als Krebsmedikamente durch die US Food and Drug Administration (FDA) zugelassen.

Diese Dissertation hat zum Ziel, neuartige Tyrosinkinase-Inhibitoren mit computergestützten Methoden (Computational Modelling) zu identifizieren. In den letzten vier Jahren wurde ein breites Spektrum an akkuraten und flexiblen Virtual Screening-Werkzeugen fertiggestellt und zusammengefasst als Mr. Dock Programm mit akkurater Gütekontrolle (Scoring Function). Etabliert wurde ausserdem ein Kontrollmechanismus per Moleküldynamik (Molecular Dynamics) um Induced Fit Problematiken, welche im Docking häufig auftreten, analysieren und lösen zu können sowie einen größeren chemischen Suchraum für die Kinase abzudecken. Desweiteren wurden In-silico fragmentbasierte Ansätze um ein neues Molekülfragmentierungsprogramm erweitert, LIBO. Die entwickelten Methoden wurden retrospektiv validiert und erzielten, prospektiv, die Identifikation von drei neuartigen Tyrosinkinase-Inhibitoren, welche in ihrer Aktivität von mikromolarer bis zu nanomolarer Effektivität variierten.

Das als erstes identifizierte pyrrolo[2,3-b] quinoxaline Gerüst (Scaffold) führte zu einer Reihe an Derivaten, welche potente enzymatische, zelluläre Aktivität, eine hohe Löslichkeit sowie ein Selektivitätsprofil, vergleichbar mit etablierten Tyrosinkinase-Inhibitoren, aufweisen. Das zweite Isoquinolin-Gerüst führte zu einem 160 nM Mustermolekül (lead) mit vielversprechenden Möglichkeiten zur Optimierung. Abschliessend wurden alle Vorhersagen durch Röntgenkristallographie (X-ray crystallography) bestätigt, was den Schluss zulässt, dass alle Entdeckungen durch die gezielte Anwendung computergestützter Methoden erreicht wurden, nicht etwa durch glückliche Zufälle.

Contents

Chapter 1

Introduction

Introduction

1.1 Role of Kinases in Diseases

Protein kinases catalyze the transfer of the γ-phosphate of ATP to protein substrates. Phosphorylation usually results in a functional change of the target protein by changing enzyme activity, cellular location, or association with other proteins. The human genome contains about 500 protein kinase genes and they constitute about 2% of all human genes. Up to 30% of all human proteins may be modified by kinase activity, and kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction.[1] Dysregulation of these pathways has been shown to be a causative factor in human diseases, making protein kinases one of the most intensively pursued classes of drug targets with about 30 distinct kinase targets being developed to the level of a Phase I clinical trial.[2] The vast majority of these targets are being investigated for the treatment of cancer. However, dysregulation of kinase function has also been implicated in other disorders, including immunological, neurological, metabolic and infectious disease. Since the first small-molecule kinase inhibitor Imatinib reached the market in 2001, 14 kinase inhibitors have received US Food and Drug Administration approvals as cancer treatments and considerable efforts are ongoing to develop selective small-molecule inhibitors for a host of other kinases that are implicated in cancer and other diseases.

1.2 Targeting Kinase Domain by Different Inhibition Mechanism

The catalytic activity of protein kinases is mediated by ATP, which binds in a deep cleft located between the N- and C-terminal lobes of a single domain. ATP binds in the cleft with the adenine ring forming hydrogen bonds with the kinase hinge, the segment connecting the N- and C-terminal lobes. The ribose and triphosphate groups of ATP bind in a hydrophilic channel extending to the substrate binding site that features conserved residues essential to catalysis. All kinases have a conserved activation loop, which is important in regulating kinase activity and is marked by conserved DFG and APE motifs at the start and end of the loop, respectively. Kinases can assume a large number of conformations with the extremes being a conformer that is catalytically competent and usually phosphorylated, and an inactive conformer in which the activation loop blocks the substrate binding site or the key catalytically residues are repositioned.

To inhibit the abnormal activity of kinases that are over expressed in cancer cells, it is then quite natural to target the ATP binding site by small molecules, which directly compete with

ATP. The majority of small-molecule kinase inhibitors developed so far target the ATP binding site of the kinase in its active state, so called type I inhibitors.[3] Given the fact that ATP binding site is highly conserved among kinases, the intellectual property (IP) landscape is highly congested and discovery of an entity chemically distinct from existing prior art becomes increasingly difficult.

The first small molecule to reach the market was imatinib (Gleevec), a so called type II tyrosine kinase inhibitor which binds to the inactive state of Abl1 characterized by a closed conformation of the activation loop (DFG-out). The flip of the DFG-motif, a conserved triad Asp-Phe-Gly at the beginning of the activation loop, induces remarkable changes in the ATP binding site and exposes an additional hydrophobic pocket that is less conserved in sequence.[4] Many type I kinase inhibitors have failed in preclinical or clinical development due to lack of selectivity causing intolerable side effects, largely because the ATP binding site is highly conserved in sequence and conformation.[5] The emergence of type II inhibitors creates new opportunities by targeting the allosteric pocket of the DFG-out conformation, offering selectivity and intellectual property novelty.[6] Bothe type I and II kinase inhibitors are ATP-competitive, but type II inhibitors further stabilize the kinase in its catalytically inactive state.

Compared to type I and II, allosteric inhibitors usually referred to as Type III inhibitors bind exclusively to sites outside of the conserved ATP site. Targeting the kinase domain outside of the conserved ATP site is regarded as to be an effective strategy to achieve greater selectivity, cellular potency and to identify inhibitors with novel IP property. The type III inhibitors can be further classified into ATP-competitive and Non-ATP-competitive. The ATP-competitive type III inhibitors bind in an allosteric pocket adjacent to the ATP site, and indirectly compete with ATP by modifying the shape of the ATP site. The Non-ATP-competitive type III inhibitors bind in an allosteric site that is distal to the ATP site, and inhibit catalysis by: 1) repositioning key catalytic residues; or 2) competing with binding of protein/peptide substrate; or 3) binding to regulatory sites or domains that are distal to the active site cleft.

1.3 Ligand-Based Virtual Screening

Ligand-based virtual screening firstly can be conducted in a rather intuitive way that is to retrieve common features (e.g., hydrogen bonding donors and acceptors) from a set of known inhibitors, and further use such information to inquire a compound library. The conventional way is to take the binding affinity as a function of physico-chemical properties, topological indices and electronic structures of ligands, and then establish a quantitative structure-activity relationship (QSAR) from a set of known inhibitors. The more complicated way is to include 3D conformers of ligands in the derivation of QSAR model.

The nature of ligand-based approach is a guess of atomic protein-ligand interaction from the sole part of ligands. When the protein structure is not available, this approach is certainly a nice choice among others, if not the best. However, the nature of guess determines its ambiguity in screening, leading to a low enrichment factor. The ambiguity in screening on the other hand could be an advantage, compared to the precision of structure-based docking approach where induced fit is a common issue limiting identification of ligands that could not fit into the rigid binding site.

1.4 Structure-Based Virtual Screening

Structure-based approach is to predict the dissociation constant on the atomic level. Starting from the thermodynamic definition of the dissociation constant, one would immediately propose to simulate the binding/unbinding of ligands from the protein by molecular dynamics (MD), and directly calculate the dissociation constant from kinetics assuming 1:1 binding:

$$K_D = k_{off} / k_{on} \qquad (1)$$

Alternatively, equipped with powerful computers, one can even propose to simulate a system with multiple copies of proteins and ligands, and directly get the dissociation constant from thermodynamic equilibrium:

$$K_D = [protein][ligand]/[complex] \qquad (2)$$

Although molecular dynamics provide a vivid image regarding how ligands bind/unbind at the atomic level, it is computationally rather expensive, and it is almost never used in practice as a virtual screening means. The MD based free energy perturbation approach such as umbrella sampling is computationally more efficient, but still seldom used in practice due to its rather slow computational speed and degraded accuracy for molecules with drug-like size.

$$\Delta G = -RT \ln(K_D) \qquad (3)$$

$$\Delta G = \Delta H - T\Delta S \qquad (4)$$

On the other hand, one can get the dissociation constant from binding free energy (Eq. 3), which is a sum of enthalpy ΔH and entropic effect $-T\Delta S$ (Eq. 4). Entropy is rather difficult to compute and quite often has been neglected, while enthalpy as interaction energy between the protein and ligand can be easily computed. In the framework of empirical force field, enthalpy is the sum of van der Waals interaction energy, electric interaction energy including the Columbic energy in vacuo and desolvation energy upon ligand binding, conformational strain of the ligand and so on. Given a binding pose of the ligand, the interaction energy can then be

easily and quickly computed, which provides an efficient means to predict the binding free energy. The primary focus for this approach is to find rational binding modes of ligands, which is called docking. The common structure-based approach consists of a docking step and a post-scoring step that utilizes a sophisticated energy function to prioritize the ligands, and both steps are usually based on a rigid protein structure.

1.5 Limitations in Current Computational Modeling

At first glance, one might be quite optimistic about drug discovery using structure-based approach, since the binding free energy can be easily computed. However, this is quite wrong if not completely. One has to realize the limitations in current structure-based approach (to name a few but not limited to):

1. The empiric force field is Newton physics based, and is approximate just as "empiric" implies. It weakly treats hydrogen bonding and pi-pi packing interaction which plays an essential role in ligand binding, and it is simply incapable of handling halogen bonding and hypercovalency of sulfur atoms.
2. Accuracy of the empiric force field is not only determined by the physical models, but also the parameters. Errors in experimental data used for parameterization have not been systematically reviewed and corrected;
3. The parameters for small molecules have not been extensively validated.
4. Most importantly, enthalpy only cannot represent the binding free energy as shown by thermodynamic data of 100 protein-ligand interactions.[7] In general, protein-ligand complexes that exhibit more negative enthalpies of binding do so at the cost of more positive $-T\Delta S$ terms and vice versa. This enthalpy-entropy compensation effect is clearly evidenced by the thermodynamic data[7] in a way that enthalpy and $-T\Delta S$ are strongly negatively correlated. Due to the absence of the entropic effect in an energy function (usually referred to as a scoring function), it inevitably leads to false positives when the entropic effects significantly deviate from those in the data set used to train the energy function. This becomes more severe for larger molecules, as the enthalpy increases with the increase in molecular size given the pairwise additive nature of van der Waals interaction while the increasing unfavorable entropic effect is missing. Although MD based approaches could in principle rigorously deal with entropic effects, computational modeling is just a means to the end for drug designers. So, a drug designer would have no reason to use any molecular dynamics based methods if they are not comparable to the high throughput screening (HTS) in terms of cost and time efficiency, not to mention that essentially all computational approaches so far are disappointing inaccurate.

1.6 Dance with Shackles on

What is the real situation of computational modeling like in current pharmaceutical industry? "The more extravagant promises of combichem ended up in the same bin as the more extravagant promises of modeling, stored in the basement alongside some dusty but snazzy-looking hardware for which no one now can locate the manual".[8] This situation is in strong contrast to the prosperous outlook in computational literature. However, one would understand the contradiction immediately if one notices that scientific journals only publish work where people managed in successful identification of active compounds, a phenomenon so called cherry pick in statistics. Serendipity always plays a certain role in the process of anything that is called discovery including drug discovery. A computational method with few successful prospective applications does not necessarily rule out the role of serendipity, and thus does not necessarily demonstrate its utility in practice.

However, there is no need to be pessimistic. It would be equally stupid to deny the importance of computational modeling, which at this moment is just not as magic as it ever promised. To understand its limitations and to complement it with our knowledge would be a simple way to be successful in drug discovery compared to the conventional HTS in terms of both time and cost, as what will exemplified below.

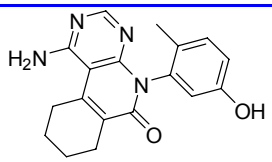
1. **Hydrogen Bonding Penalty**: Hydrogen bonding is one of the most appreciated concepts in the field of drug design, but could also be most superficially appreciated in a way that most people simply prefer more and stronger hydrogen bonds. However, experiment has shown that mutation of Cys-35 in Tyrosyl-tRNA synthetase (TyrTs) to Ser-35 caused poorer ATP binding and catalysis although the hydroxyl group of serine forms far stronger hydrogen bonds than does the thiol group of cysteine.[9] In Chapter 2, change in hydrogen bonds during the binding process was reviewed. In an aqueous environment, formation of hydrogen bonds between ligand and receptor inevitably is accompanied by breakage of hydrogen bonds with water. The enthalpic gain by formation of hydrogen bonds is marginal though it is entropically favorable by releasing constrained water into the bulk. From the perspective of enthalpy, the formation of hydrogen bonds between ligand and receptor does no good, but leads to a potential enthalpic loss if the newly formed hydrogen bonds are not in optimal geometry. A scoring function with enthalpic penalty in hydrogen bonding was presented, leading to the discovery of a novel class of potent, selective and soluble tyrosine kinase inhibitors currently under investigation towards clinical application.
2. **An Induced Fit**: The phenomenon of induced fit where the ligand binding site alters upon binding to become very specific for the given ligand limits the use of the crystal

structure in virtual screening campaigns. To circumvent this issue and identify inhibitors of tyrosine kinases in their inactive form (DFG-out), a molecular dynamics approach was taken in Chapter 3. The computation approach identified a novel chemotype for type II kinase inhibitors, which also inhibits the therapeutically relevant T315I mutant of Abl1.

3. **In Silico Fragment-Based Approach**: Fragment-based drug design approach has seen many successful applications in a cost and time efficient way. In Chapter 4, a computationally efficient algorithm was introduced to decompose a compound library, in order to obtain anchor fragments with high chemical richness that can serve as a starting point either directly for hit optimization or for identification of their "parent" compounds. The prospective application was illustrated on EphB4 kinase, leading to the discovery of a novel chemotype of tyrosine kinase inhibitors. This *in silico* fragment based approach is able to screen the entire compound library in a very efficient way!
4. **Docking**: Docking, as computational simulation of a candidate ligand binding to a receptor, is used to predict the binding poses of small-molecule drug candidates into their protein targets of pharmaceutical relevance, and is the key in structure-based virtual screening. In Chapter 5, a new docking software Mr. Dock, standing for Molecular Recognition in flexible Docking of small molecules to proteins, was presented, and completes our virtual screening "tool box".
5. **Discovery of Novel ZAP70 Inhibitors**: With the complete "tool box", virtual screening now becomes a routine yet boring job, but identification of active pharmaceutical ingredients is still struggling.
6. **Deterministic or Serendipity?** Is a discovery deterministic or serendipity? This is always arguable but certainly a good question to think about. A great discovery is still great no matter how it is made, but a computational scientist should care more about the general applicability of the methods with careful evaluation of the role of serendipity in the discovery. In Chapter 7, the identification of novel EphB4 inhibitors is used as an example to initiate the discussion.

1.7 Highlights of Discovered Lead Compounds

Table 1. Lead Compounds[a] Derived from Indentified Hits.

Internal Code	Structure	Cellular activity on EphB4 (nM)	Kinetics[b]	
			K_D (nM)	$1/k_{off}$ (min)
ln06		930	ND[c]	ND
lm10	undisclosed	<6	6.5	0.24
ln28	undisclosed	20	35	25
ln38	undisclosed	18	9.4	157

[a]Synthesized by Dr. Karine Lafleur. [b]By surface plasma resonance. [c]Not determined.

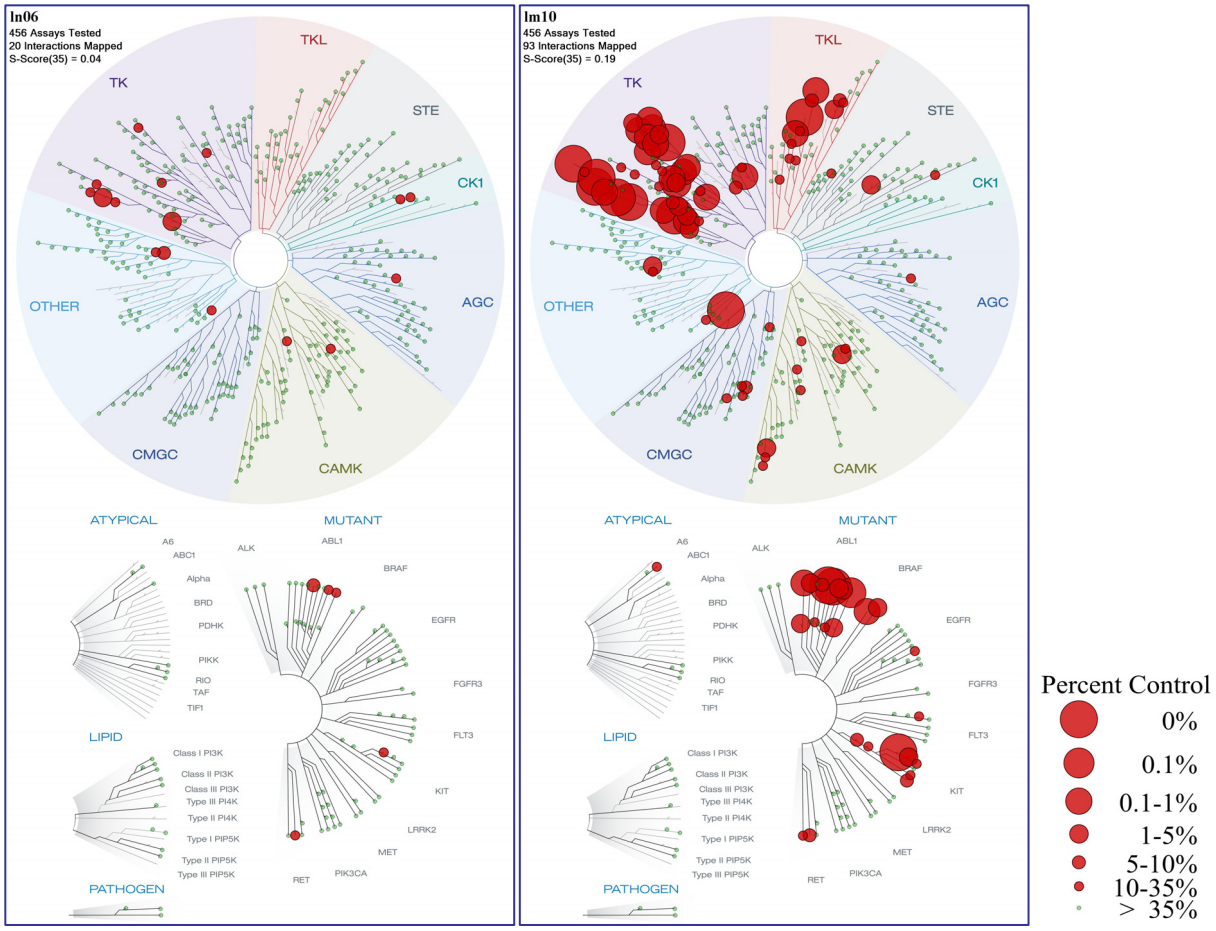


Figure 1. Selectivity profile of lead compounds ln06 and lm10 assayed at 1 μM (small values of big red circles indicate strong inhibition).

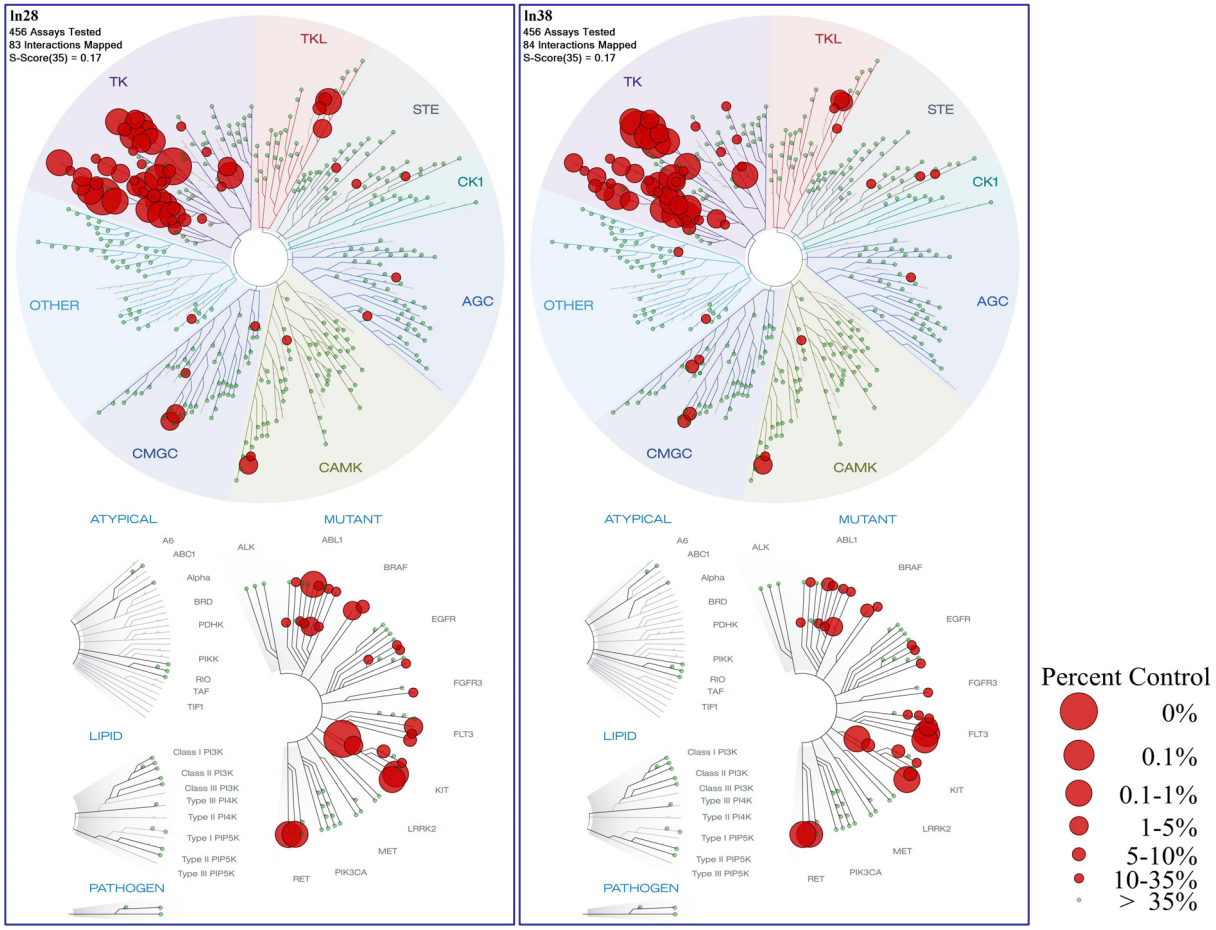


Figure 2. Selectivity profile of lead compounds ln28 and ln38 assayed at 1 µM (small values of big red circles indicate strong inhibition).

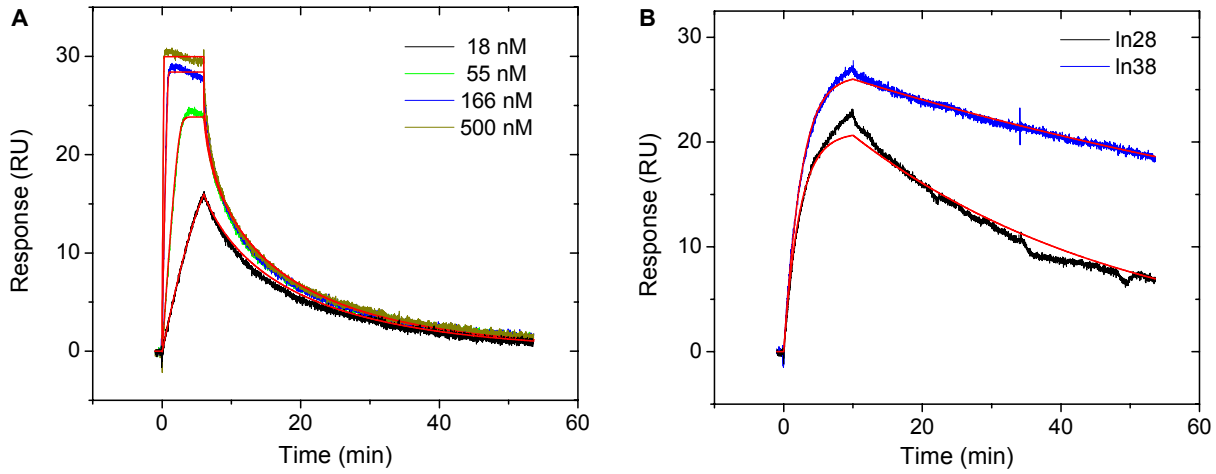


Figure 3. Inhibitor binding to EphA3 surfaces. Biacore response data for lm10 at four concentrations (A) and for ln28 and ln38 both at a concentration of 500 nM (B), along with kinetic curve fits in red.

Reference

1. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002,** 298, 1912-34.
2. Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* **2009,** 9, 28-39.
3. Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2006,** 2, 358-64.
4. Noble, M. E.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004,** 303, 1800-5.
5. Liao, J. J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* **2007,** 50, 409-24.
6. Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *J Med Chem* **2010,** 53, 2681-94.
7. Reynolds, C. H.; Holloway, M. K. Thermodynamics of Ligand Binding and Efficiency. *Acs Med Chem Lett* **2011,** 2, 433-437.
8. Lowe, D. B. Nowhere To Go But Up: The Return of Medicinal Chemistry. *Acs Med Chem Lett* **2012,** 3, 3-4.
9. Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Winter, G. Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. *Biochemistry* **1983,** 22, 3581-6.

Chapter 2

Hydrogen Bonding Penalty upon Ligand Binding

OPEN ACCESS Freely available online PLoS one

Hydrogen Bonding Penalty upon Ligand Binding

Hongtao Zhao, Danzhi Huang*

Department of Biochemistry, University of Zurich, Zurich, Switzerland

Abstract

Ligand binding involves breakage of hydrogen bonds with water molecules and formation of new hydrogen bonds between protein and ligand. In this work, the change of hydrogen bonding energy in the binding process, namely hydrogen bonding penalty, is evaluated with a new method. The hydrogen bonding penalty can not only be used to filter unrealistic poses in docking, but also improve the accuracy of binding energy calculation. A new model integrated with hydrogen bonding penalty for free energy calculation gives a root mean square error of 0.7 kcal/mol on 74 inhibitors in the training set and of 1.1 kcal/mol on 64 inhibitors in the test set. Moreover, an application of hydrogen bonding penalty into a high throughput docking campaign for EphB4 inhibitors is presented, and remarkably, three novel scaffolds are discovered out of seven tested. The binding affinity and ligand efficiency of the most potent compound is about 300 nM and 0.35 kcal/mol per non-hydrogen atom, respectively.

Citation: Zhao H, Huang D (2011) Hydrogen Bonding Penalty upon Ligand Binding. PLoS ONE 6(6): e19923. doi:10.1371/journal.pone.0019923

Editor: Peter Butko, Anne Arundel Community College, United States of America

Received March 23, 2011; **Accepted** April 13, 2011; **Published** June 17, 2011

Funding: This work was supported by a grant (31003A_122442) of the Swiss National Science Foundation (www.snf.ch) to D.H. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dhuang@bioc.uzh.ch

Introduction

Hydrogen bonding is an exchange reaction whereby the hydrogen bond donors and acceptors of the free protein and ligand break their hydrogen bonds with water and form new ones in the protein-ligand complex [1,2,3]. About thirty years ago, Wilkinson and coworkers found mutation of Cys-35 in Tyrosyl-tRNA synthetase to Ser-35 causes poorer ATP binding and catalysis although the hydroxyl group of serine forms far stronger hydrogen bonds than does the thiol group of cysteine [1]. Analysis of the hydrogen bonding geometry revealed that a hydrogen bond of Ser-35 is at least 0.5 Å longer than the optimum. Accordingly, Ser-35 would have to lose a good hydrogen bond with a bound water molecule to form this weak hydrogen bond with ATP in the enzyme-substrate complex, and thus the mutant shows poorer binding and catalysis. Therefore, enthalpic loss in hydrogen bonding could take place upon ligand binding if not compensated by formation of good hydrogen bonds between the protein and ligand.

Virtual screening has emerged as an efficient tool in drug discovery from lead identification to optimization and beyond [4,5]. However, scoring functions that model the solvent environment as a continuum [6,7] are still grossly inaccurate [8]. The role of individual waters can be critical in predication of binding affinities, and continuum models often provide poor results in treating bound waters in a confined cavity [9]. Glide docks explicit waters into the binding site and measures the exposure of polar/charged groups to the explicit waters. When a polar/charged ligand or protein group is judged to be inadequately solvated, a desolvation penalty is assessed [9,10]. By contrast, most other scoring functions [11] do not properly take into account the enthalpic loss of hydrogen bonding upon ligand binding. Incorporation of bound water molecules into molecular docking was suggested for improvement of accuracy [12]. On the other hand, in high-throughput molecular docking campaigns a significant part of binding poses are rather unrealistic, e.g. burial of polar atoms in hydrophobic sites, and thus discarding them at an early stage is desirable. Filters such as van der Waals efficiency based on arbitrary cutoff are often used to remove poses that unlikely bind [13]. However, it seems lack of a reliable and efficient filter with transferable cutoff among different proteins.

Protein kinases play an important role in cell-signaling pathways regulating a variety of cellular functions. Dysregulation of kinase activity has been implicated in pathological conditions ranging from neuronal disorders to cellular transformation in leukemia [14]. The tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4) is involved in cancer related angiogenesis [15]. So far, two high-throughput virtual screening campaigns have been reported, with two scaffolds identified in the low micromolar range [13,16]. Highly potent EphB4 inhibitors have been developed via chemical synthesis [17,18,19]. The marketed drug dasatinib, with Abl1 and Src as primary targets, also shows a very high affinity to Eph kinases [20].

Here, we report a new approach to calculate hydrogen bonding penalty (HBP) associated with ligand binding. HBP is further integrated into a binding energy calculation, and the fitted parameter of 1.7 kcal/mol is consistent with the estimate of contribution by formation of one neutral hydrogen bond ranging from 0.5 to 1.5 kcal/mol [21]. Moreover, statistics of HBP in kinase crystal structures and an application in a high-throughput docking campaign is presented.

Methods

Binding of a ligand to a protein involves the breakage of hydrogen bonds with water molecules and formation of new hydrogen bonds between the protein and ligand, which can be described by the following equation [21] by using one pair of

donor (D) and acceptor (A):

$$D\cdots OH_2 + A\cdots HOH \leftrightarrow D\cdots A + HOH\cdots OH_2 \quad (1)$$

Based on hydrogen bonding being an exchange reaction [1,21,22], its energy can be described using normalized weights:

$$E_{HB-unbound} = [(w_D + w_{O-H_2O}) + (w_A + w_{H-H_2O})] * E_{HB} \quad (2)$$

$$E_{HB-bound} = [f_{hb} * (w_D + w_A) + (w_{O-H_2O} + w_{H-H_2O})] * E_{HB} \quad (3)$$

wherein, w_D and w_A is the hydrogen bonding weight of a donor or acceptor, respectively, f_{hb} stands for the fraction of hydrogen bonding relative to that of an optimum geometry, and E_{HB} is unit hydrogen bonding energy. Hydrogen bonds with water are assumed to be in the optimum geometry. HBP (p_{HB}) associates with ligand binding is then described as

$$p_{HB} = (1 - f_{hb}) * (w_D + w_A) \quad (4)$$

Probing hydrogen bonding status

Oxygen and nitrogen atoms in double or triple bonds are regarded as hydrogen bond acceptors, and hydrogen atoms bonded to oxygen, nitrogen or sulfur atoms are regarded as hydrogen bond donors. The existence of C—H...O hydrogen bonds has been confirmed by neutron diffraction data on organic compounds [23]. Analysis of 100 kinase crystal structures complexed with small molecule inhibitors at a resolution of at least 2.5 Å gives 64 short C—H...O interactions, showing typical hydrogen bonding features (Figure S1).

Each hydrogen bond donor or acceptor at the binding interface is firstly checked whether it forms hydrogen bond with water molecules. For this purpose, an optimum solvation radius (r_{sol}) is defined for each donor/acceptor and if a water molecule can be placed within 0.15 Å of the r_{sol} no penalty is applied. Here, 2.8 and 2.9 Å are used as r_{sol} for any oxygen and nitrogen, respectively, which were derived from an analysis of 397 crystal structures with X-ray resolutions below 1.0 Å (Figure S2). The r_{sol} of polar hydrogen is 1.9 Å (except 2.15 Å for H bonded to sulfur), which is the difference between the r_{sol} of nitrogen and the bond length [24] of N—H. The r_{sol} of other atom types are listed in Figure 1 and the values are mainly adapted based on the van der Waals radii of Bondi [25]. Details of probing hydrogen bonds with water were described in File S1. In case of not forming hydrogen bonds with water, the possibility of forming hydrogen bonds between the protein and ligand (including intra-molecular hydrogen bonds) is further checked and penalty (p_{HB}) is then calculated.

Fraction of hydrogen bonding

Similar to the strategy of evaluating hydrogen bonding energy in LUDI [26], the following equations are used to calculate the fraction of hydrogen bonding (f_{hb}) to that of an optimum geometry.

$$f_{hb}(r,\theta) = f(r)\cdot f(\theta) \quad (5)$$

	Atom types	Hydrogen bonding weight	r_{sol} (Å)
O	=O, $-O^-$	1.0	2.8
	−OH	0.5	
	−O−	0.2	
N	=N−, ≡N	1.0	2.9
H	$-\overset{\overset{\mathrm{H}}{\mid}}{\mathrm{N}}H^+$	1.0	1.88
	−OH, $-\overset{\overset{\mathrm{H}}{\mid}}{\mathrm{N}}H$, $-NH_3^+$, $-\overset{\overset{\mathrm{H}}{\mid}}{\mathrm{N}}H_2^+$	0.5	
	$-NH_2$	0.3	
	CH=N, HC=O	0	
	−SH	0.2	2.15
S	=S, −S−	-	3.2
C	Unsaturated	-	3.0
	Saturated	-	3.2
H	Apolar	-	2.28
F	-	-	2.99
Cl	-	-	3.27
Br	-	-	3.37
I	-	-	3.50

Figure 1. Hydrogen bonding weights and solvation radii of different atom types.
doi:10.1371/journal.pone.0019923.g001

$$f(r) = \begin{cases} 1 & if \quad r \le 2.0 \\ 1 - 0.5 * (r - 2.0) & if \quad 2.0 < r \le 2.8 \\ 0 & if \quad r > 2.8 \end{cases} \quad (6)$$

$$f(\theta) = \begin{cases} 1 & if \quad \theta \ge 150^o \\ 4.0 * (-\cos\theta)^{0.25} - 2.86 & if \quad 110^o \le \theta < 150^o \\ 0 & if \quad \theta < 110^o \end{cases} \quad (7)$$

wherein, r is the distance between the hydrogen atom and the acceptor and θ is the angle centered at hydrogen among donor, hydrogen and acceptor. The equation to calculate $f(r)$ and $f(\theta)$ as well as the upper and lower limit in r and θ are derived from the calculation using density functional theory [27]. In case of one hydrogen atom is shared by two acceptors or one acceptor interacting with two donors, the f_{hb} for the corresponding donor/acceptor is additive but with 1 as the upper limit.

Hydrogen bonding penalty

The HBP at the protein-ligand interface is summarized over each donor/acceptor as

$$P_{HB} = \sum_{pro,lig} w * (1 - \sum f_{hb}) \quad (8)$$

However, no penalty is applied for protein atoms which are not water accessible before ligand binding or participate in intramolecular hydrogen bonds. Initial guess of hydrogen bonding weights (w) is based on chemical intuition by considering atomic partial charge and water solubility of a few small molecules (Table S1). Empirical weights as proof-of-principle are then optimized with a trial-and-error procedure according to the fitted parameter in the binding free energy calibration.

Evaluation of binding free energy

The equation used for fitting the calculated energies to the experimental free energies of binding ($\Delta G = RT\ln(K_d)$) is a three-parameter model

$$\Delta G = \alpha \Delta E_{ff} + \beta P_{HB} + \gamma \quad (9)$$

where, ΔE_{ff} is the interaction energy between the ligand and the protein calculated by the CHARMm force filed [28] and P_{HB} stands for HBP. Three parameters α, β, and γ are generated with fitting. ΔE_{ff} is calculated by the following equation:

$$\Delta E_{ff} = \Delta E_{vdW} + \Delta E_{coul} + \Delta G_{solv} + \Delta E_{strain} \quad (10)$$

where, ΔE_{vdW} is the intermolecular van der Waals energy, ΔE_{coul} is the intermolecular Coulombic energy in vacuo, $\Delta E_{\mathrm{strain}}$ is the strain energy of ligand upon binding, and ΔG_{sol} is the change in solvation energy of ligand and protein upon binding.

The van der Waals and Coulombic interaction energy are calculated by subtracting the values of the isolated components from the energy of the complex with CHARMM [29] and the CHARMm22 force filed [28]. The van der Waals energy is calculated using the default nonbonding cutoff of 14 Å. Coulombic energy is calculated using infinite cutoff and a dielectric constant of 2.0. The electrostatic solvation energy was calculated by the finite-difference Poisson approach (FDP) [30] using PBEQ module [31] in CHARMM and a focusing procedure with a final grid spacing of 0.25 Å. The size of the initial grid is determined by considering a layer of at least 12.5 Å around the solute. The dielectric discontinuity surface was delimited by the van der Waals surface. The ionic strength is set to zero and the temperature to 300 K. Two finite-difference Poisson calculations are performed for each of the three systems (protein, ligand, and protein/ligand complex). The exterior dielectric constant was set to 78.5 and 2.0 for the first and second calculation, respectively, while the solute dielectric constant is 2.0 to take polar fluctuations into account. The solvation energy is the difference between the two calculations. The strain energy of the ligand is the energy difference between the bound and global minimum. Here, the global minimum is the one showing the lowest $E_{\mathrm{vdW}}+E_{\mathrm{coul}}+E_{bonded}+G_{\mathrm{sol}}$ among all the poses that have been minimized outside of the protein.

Twenty-three inhibitors [32] of CDK2 (1H0V), 24 inhibitors [18] (**8** to **32**, excluding **30**) of EphB4 (2VWX), and 27 uncharged inhibitors [33] of p38 alpha MAP kinase (3GC7) are used as the training set. Thirty type II inhibitors [34] of Braf (3II5), 14 charged inhibitors [33] of p38 alpha and another 20 p38 alpha inhibitors [35] (1YWR) are used as the test set. Protein structures were taken from the X-ray structure (PDB code indicated in the brackets) and prepared as described below. Some key physiochemical properties of inhibitors are summarized in Figure S3.

Version 4 of AutoDock [36] was used to generate the binding poses over the conformational search space using the Lamarckian genetic algorithm. The binding site was determined by 4.0 Å away from any atom of the ligand complexed in the respective protein structure. The number of energy evaluations was 2,750,000 and the number of poses was 50. Poses were further clustered using all atom RMSD cutoff of 0.3 Å to remove redundancy and in average 20 cluster representatives were kept. All other parameters were set as default. A few poses for each inhibitor were also generated by manual modification of the scaffold present in the respective crystal structure. All poses were further minimized by CHARMM in the respective proteins. The protein structure was kept rigid in all steps.

Preparation of protein-ligand complexes

One hundred kinase crystal structures (including 15 different classes, File S2) complexed with small molecule inhibitors at a resolution of at least 2.5 Å were downloaded from Protein Data Bank for analysis of HBP. Hydrogen atoms were added according to the protonation states of chemical groups at pH 7. Partial charges were then assigned using MPEOE method [37,38]. The added hydrogen atoms were minimized by the conjugate gradient algorithm to a RMS of the energy gradient of 0.01 kcal $\cdot$ mol^{-1} Å^{-1}. During minimization, the electrostatic energy term was screened by a distance-dependent dielectric of $4r$ to prevent artificial deviations due to vacuum effects, and the default nonbonding cutoff of 14 Å was used. Furthermore, the positions of all heavy atoms were fixed.

Preparation of the compounds library for virtual screening

The compounds were selected from Zinc library [39]. Preparation included the assignment of CHARMm atom types, force field parameters [28], and partial charges [37,38], and energy minimization with a distance dependent dielectric function using the program CHARMM [29].

Enzymatic assay

In vitro kinase activity was measured using the Panvera Z'lyte Tyr2 kinase assay PV3191 (Invitrogen) according to the

manufacturer's instructions. The reaction assay (10 μL) contained 7.5 ng of EphB4 kinase (Proqinase, Germany), 30 μM ATP, and 5% DMSO. The reaction was performed at room temperature for 1 h.

Results and Discussion

Statistics of hydrogen bonding penalty in kinase complexes

Small HBPs can be observed for the binding modes of inhibitors in the X-ray structures. One example is c-Kit tyrosine kinase with its apo and holo form in complex with Imatinib (PDB codes 1T45 and 1T46). In the apo conformation, donors/acceptors at the ATP binding site form hydrogen bonds with bound water molecules. While upon ligand binding, as shown in the holo conformation, some water molecules are displaced by Imatinib. HBP on the protein part is close to zero because new hydrogen bonds to the protein are formed to compensate for the replacement of the water molecules. However, one nitrogen atom of the Imatinib pyrimidine ring (N1 of Figure S4) becomes water inaccessible and does not form a new hydrogen bond, leading to a penalty of 1. By contrast, the other nitrogen atom (N2 of Figure S4) remains hydrogen bonding with a nearby bound water molecule and thus has no penalty.

To check the distribution of HBP values in crystal structures, 100 kinase-ligand complexes are investigated. In this data set, all the small molecule inhibitors have molecular weights from 200 to 700 g/mol and number of donors or acceptors from 2 to 11 (File S2). The HBP has been calculated for each of them and the values are in general small, with 62% smaller than 1 and 36% and 2% in the range from 1 to 2 and 2.0 to 2.1, respectively (Figure 2 and File S2). It has also been observed that larger HBPs appear in some X-ray structures, e.g., the structures of PDB code 3KVX and 1JSV, and the large values actually originate from poor fitting of small molecules to the density, a common problem in crystallography [40] which can be manifested by clash of atoms.

Distribution of HBPs for docked poses of small-molecule inhibitors is also evaluated. Here, the 138 molecules used in the binding free energy calibration are docked into the corresponding protein binding sites with AutoDock. For each molecule, about 20 poses in average are generated. Then the HBPs and binding energies are calculated for all the poses. Firstly, the binding pose with the most favorable binding energy for each molecule (Figure S5) is selected and the distribution of HBPs is plotted. As can be observed from **B** of Figure 2, the distribution is similar to that of the 100 kinase complex structures (**A**). On the other hand, the distribution of all poses (**C**) spreads more widely with the largest HBP being 6.5. Compared with the HBPs in the crystal structures (**A**), 2 is a reasonable threshold, and about 50% of poses with unrealistic binding modes can be filtered out from further evaluations.

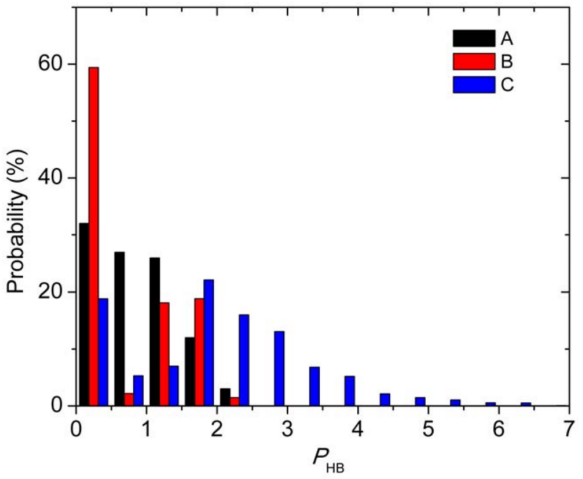


Figure 2. Distribution of hydrogen bonding penalties for: A) the binding modes in crystal structures of the 100 kinase complexes; B) poses with the most favorable calculated binding energies of the 138 molecules used in binding free energy calibration; C) all poses of the 138 molecules.
doi:10.1371/journal.pone.0019923.g002

Hydrogen bonding penalty improves the accuracy of binding energies calculation

Binding energies can be calculated using equation 9 with the parameters obtained by least-squares fitting on the training data

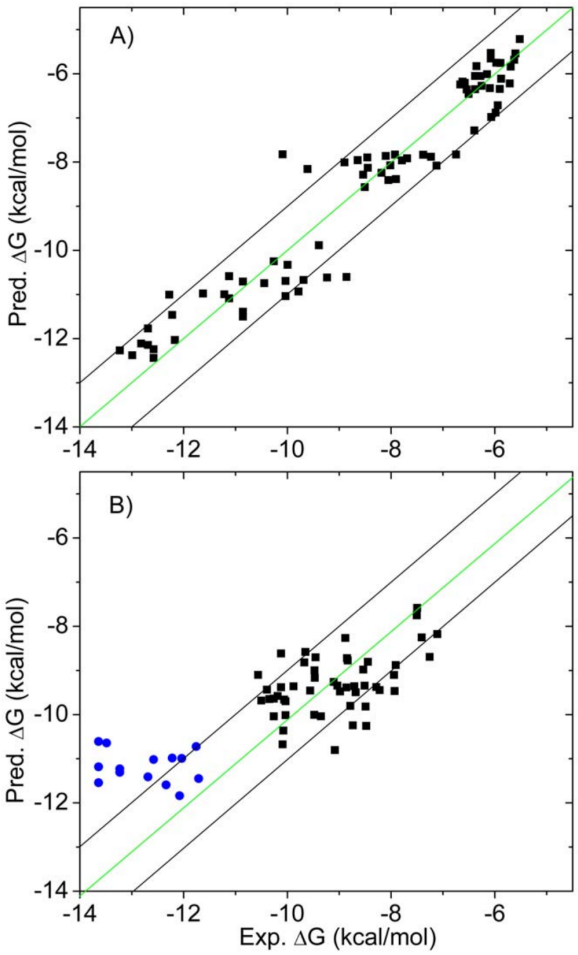


Figure 3. Comparison of the calculated versus experimental binding energies. A) Training set of 74 inhibitors. $R^2 = 0.92$ and RMS error = 0.69 kcal/mol; B) Validation set of 64 inhibitors. RMS error = 1.12 kcal/mol. The blue dots indicated the 14 p38α inhibitors with one formal charge. The green diagonal line is the ideal line of perfect prediction. The black diagonals delimit the 1 kcal/mol error region.
doi:10.1371/journal.pone.0019923.g003

Table 1. Further validation of the three-parameter model with kinases and aspartic protease.

Protein	PDB code	ΔE_{ff} (kcal/mol)	P_{HB}	ΔG_{pred} (kcal/mol)	ΔG_{exp} (kcal/mol)
Abl	1OPJ	−64.80	1.24	−12.45	−10.81
Braf	1UWH	−57.61	1.27	−10.91	−10.45
JAK2	3E63	−30.18	0.00	−7.41	−7.91
Lck	2OFV	−59.13	0.53	−12.51	−13.23
JNK3	1PMV	−30.16	0.17	−7.12	−9.31
Ret	2X2L	−26.67	0.07	−6.58	−7.20
EGFR	1XKK	−66.60	2.30	−11.00	−10.91
CSrc	3G5D	−52.34	1.64	−9.19	−12.82
HIV-1 protease	1HIH	−65.71	1.49	−12.21	−11.01
	1HPX	−65.44	1.43	−12.26	−12.46
	1HXB	−61.24	0.95	−12.21	−13.49
	1HXW	−72.66	1.41	−13.78	−14.71
BACE-1	2QMF	−73.62	1.56	−13.72	−11.63
	2QP8	−68.76	0.47	−14.59	−11.05
	2XFI	−71.10	2.36	−11.83	−10.67

doi:10.1371/journal.pone.0019923.t001

set of the 74 CDK2, EphB4, and p38α inhibitors as following:

$$\Delta G = 0.207 * \Delta E_{ff} + 1.72 * P_{HB} - 1.17 \qquad (11)$$

The calculated binding energies show high correlation with the experimental values (R-square of 0.92) and a small RMS error of 0.69 kcal/mol (Figure 3A). Here, the parameter β corresponds to the unit hydrogen bonding energy. Notably, the fitted value 1.72 kcal/mol is in agreement with the experimental value, e.g., breakage of a neutral hydrogen bond resulting in loss of energy from 0.5 to 1.5 kcal/mol [21]. Moreover, a charged primary amine or carboxyl group has a hydrogen bonding weight of 1.5 or 2.0, which can lead to a maximal penalty of 2.58 or 3.44 kcal/mol upon loss of the hydrogen bond/salt bridge. This value also agrees well with the experimental data (up to 4 kcal/mol) [21]. Hydrogen bonding weights were further used to rank the strength of individual hydrogen bonds in DNA base pairs, exhibiting good compatibility with the previously reported results (File S3).

The fitted model has been validated on a test set including 14 charged p38α inhibitors and 30 type II Braf inhibitors, with an RMS error of 1.12 kcal/mol (Figure 3B). Moreover, validation with different kinases shows general transferability of this model (Table 1). Transferability can be also seen for aspartic protease, e.g., HIV-1 protease and β-secretase, although a shift of 2.0 kcal/mol can be observed for the latter. Previously, we reported a two-parameter LIECE model for kinase inhibitors [13], which is not transferable for type II kinase inhibitors, HIV-protease or β-secretase inhibitors. The binding affinities predicted by the two-parameter LIECE on the 24 type I EphB4 inhibitors show about −5.0 kcal/mol shift compared with the experimental values (Table S2). Clearly, the incorporation of HBP into the scoring function improves the general transferability besides the role of ligand reorganization energy [41].

The derived model includes calculation of solvation energy by FDP which requires about 6 min on a single Intel 2.8 GHz CPU. Replacing the FDP approach with a distance-dependent dielectric model for solvation energy calculation gives similar accuracy for the neutral inhibitors at a much fast speed (10 seconds). However, distance-dependent dielectric model can only apply for non-charged compounds due to inaccurate treatment of the solvation effect, and also more false positives in a high-throughput virtual screening are observed. This comparison indicates that accurate calculation of solvation energies in prediction of binding affinities is necessary.

Virtual screening for EphB4 inhibitors

In a recent high throughput docking study for EphB4 inhibitors, ZINC "leads-now" library of about 20 million compounds

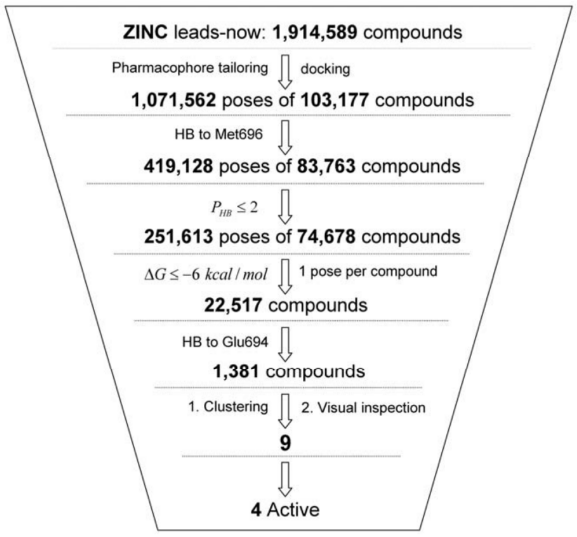


Figure 4. Schematic picture of the high throughput docking approach. HB stands for hydrogen bond. Met696 and Glu694 are the two key residues of the hinge loop (see also Figure 6).
doi:10.1371/journal.pone.0019923.g004

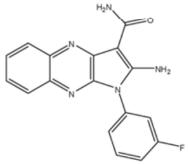
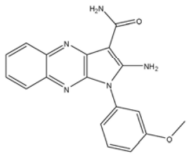
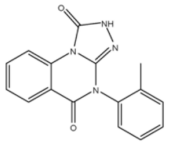
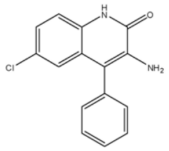
	Compound	Mw (g/mol)	P_{HB}	Pred. ΔG (kcal/mol)	IC_{50} (μM)[a]
1		321	0.22	-7.9	0.38
2		333	0.74	-8.0	0.30
3		292	1.06	-6.6	15.7
4		270	0.61	-6.4	42% at 20 μM

Figure 5. Identified EphB4 inhibitors by high throughput docking. [a] All IC_{50} values are means of two to four dose-response measurements.
doi:10.1371/journal.pone.0019923.g005

(Mw≤350 and cLogP≤3.5) was first tailored by a pharmacophore model to generate a focused library of 103,177 compounds. This pharmacophore model was specifically designed for EphB4 type I inhibitors, consisting of a bi-dentate hydrogen bonding pattern and a conjugate hydrophobic group to be located in the deep ATP back pocket as well as geometric constraints thereof (H. Zhao, unpublished results). To our best knowledge, all known type I EphB4 inhibitors [13,16,17,18] can fulfill this model.

The focused library was docked by AutoDock 4 and about 1 million poses were generated by clustering with a RMSD cutoff of 1.0 Å. The cluster representatives which do not form a hydrogen bond to NH of Met696 were further filtered out. The HBP (≤2) was then used to remove unrealistic poses (about 40%). The remaining poses were further ranked by the predicted binding energy, and the top about 30% compounds (22,517) with calculated binding energy smaller than −6 kcal/mol (~50 μM) were kept. Among them, 1381 compounds forming a hydrogen bond to Glu694 were selected and can be classified into 80 structural scaffolds. Finally, 7 scaffolds (9 compounds) of them were purchased for experimental measurements based on visual inspection of the binding modes, commercial availability and structural novelty. The procedures used in the virtual screening are shown in Figure 4. Comparison of the performances between the proposed and AutoDock 4 scoring function is shown in Figure S6.

Notably, 4 of the 9 tested compounds show inhibitory activity at micro-molar to high nano-molar range, with the most active compound showing IC_{50} at 300 nM (Figure 5). Interestingly, the two compound also show a high ligand efficiency [42] of −0.35 kcal/mol per non-hydrogen atom. The predicted binding mode of compound **3** (Figure 6) is further confirmed by the preliminary X-ray crystallography (J. Dong, unpublished results).

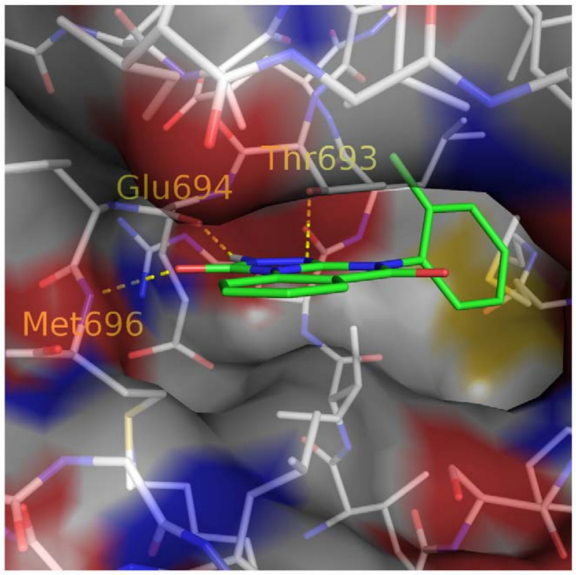


Figure 6. Binding mode of compound 3 (carbon atoms in green) predicted by docking. The intermolecular hydrogen bonds to the residues at the hinge loop (Glu694 and Met696) and the gatekeeper (Thr693) are shown by yellow dashed lines. The protein surface is colored based on atom types with carbon in white, oxygen in red, and nitrogen in blue. This figure was prepared using PyMOL (Delano Scientific, San Carlos, CA).
doi:10.1371/journal.pone.0019923.g006

Conclusion

Hydrogen bonding in biological system is a complex phenomenon as water competes with ligand for the hydrogen bonding sites. Removal of a group that forms a hydrogen bond in unfavorable geometry actually improves binding [21]. In view of hydrogen bonding being an exchange reaction [1,21,22], a new approach is proposed to evaluate the HBP upon ligand binding. Analysis of the 100 crystal structures indicates the penalty in general is low, predominantly smaller than 2 for inhibitors. A high throughput docking case shows HBP can function as an efficient filter to remove poses that unlikely bind. Incorporation of HBP into binding free energy calculation can significantly improve the predictive accuracy and transferability. The fitted parameter of 1.72 kcal/mol means loss of a neutral hydrogen bond would result in a penalty of from 0.34 to 1.72 kcal/mol in binding energy, consistent with the experimental data from 0.5 to 1.5 kcal/mol [21]. Four inhibitors of three scaffolds were discovered out of nine tested, and the binding affinity and ligand efficiency of the most potent compound is about 300 nM and 0.35 kcal/mol per non-hydrogen atom, respectively.

Supporting Information

Figure S1 Scatter plot of C—H…O angles against H…O distances in short C—H…O interactions between ligands and proteins.
(DOC)

Figure S2 Distribution of distances between crystal water oxygen and oxygen or nitrogen atoms of proteins.
(DOC)

Figure S3 Distribution of some key properties of the inhibitors used in the training and test set.
(DOC)

Figure S4 2D plot of the binding mode of Imatinib. Upon ligand binding, one nitrogen atom of the Imatinib pyrimidine ring (N1) becomes water inaccessible and does not form a new hydrogen bond, leading to a penalty of 1. By contrast, the other nitrogen atom (N2) remains hydrogen bonding with a nearby bound water molecule and thus has no penalty.
(DOC)

Figure S5 Poses with the most favorable binding energy of inhibitors of CDK2 (A), EphB4 (B), p38 α (C), Braf (D) and another set of p38α inhibitors (E). The molecules with bonds in red are the binding modes of the corresponding scaffolds in the crystal structures.
(DOC)

Figure S6 Distribution of predicted binding affinities by Autodock4 (black) and the proposed scoring function (red) on 74,678 compounds passing the first two filters (HB to Met696 and $P_{HB} \leq 2$ kcal/mol). Bin size: 0.1 kcal/mol.
(DOC)

Table S1 MPEOE partial charge and water solubility of model small molecules used to generate initial guess of hydrogen bonding weights.
(DOC)

Table S2 Two-parameter LIECE energy and hydrogen bonding penalty on the 24 EphB4 inhibitors.
(DOC)

File S1 Probing hydrogen bonds formed with implicit water.
(DOC)

File S2 Hydrogen bonding penalty of the 100 kinase complex structures.
(DOC)

File S3 Ranking the strength of individual hydrogen bonds in DNA base pairs.
(DOC)

Acknowledgments

We thank Dr. Amedeo Caflisch for useful discussions and comments on the manuscript. We thank Dr. Jing Dong for the preliminary X-ray structure. We are grateful to Armin Widmer (Novartis Basel) for continuous support with the program WITNOTP, which was used for visual analysis. Calculations were performed on the Schroedinger cluster at the Informatikdienste, University of Zurich.

Author Contributions

Conceived and designed the experiments: DH HZ. Performed the experiments: HZ. Analyzed the data: HZ DH. Contributed reagents/materials/analysis tools: HZ. Wrote the paper: HZ DH.

References

1. Wilkinson AJ, Fersht AR, Blow DM, Winter G (1983) Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. Biochemistry 22: 3581–3586.
2. Winter G, Fersht AR, Wilkinson AJ, Zoller M, Smith M (1982) Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. Nature 299: 756–758.
3. Wilkinson AJ, Fersht AR, Blow DM, Carter P, Winter G (1984) A large increase in enzyme-substrate affinity by protein engineering. Nature 307: 187–188.
4. Bajorath J (2002) Integration of virtual and high-throughput screening. Nat Rev Drug Discov 1: 882–894.
5. Langer T, Hoffmann RD (2001) Virtual screening: an effective tool for lead structure discovery? Curr Pharm Des 7: 509–527.
6. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268: 1144–1149.
7. Feig M, Onufriev A, Lee MS, Im W, Case DA, et al. (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25: 265–284.
8. Schneider G (2010) Virtual screening: an endless staircase? Nat Rev Drug Discov 9: 273–276.
9. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, et al. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49: 6177–6196.
10. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47: 1739–1749.
11. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46: 2287–2303.
12. Thilagavathi R, Mancera RL (2010) Ligand-protein cross-docking with water molecules. J Chem Inf Model 50: 415–421.
13. Kolb P, Huang D, Dey F, Caflisch A (2008) Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. J Med Chem 51: 1179–1188.
14. Hunter T (1998) The role of tyrosine phosphorylation in cell growth and disease. Harvey Lect 94: 81–119.
15. Pennisi A, Ling W, Li X, Khan S, Shaughnessy JD, Jr., et al. (2009) The ephrinB2/EphB4 axis is dysregulated in osteoprogenitors from myeloma patients and its activation affects myeloma bone disease and tumor growth. Blood 114: 1803–1812.
16. Zhou T, Caflisch A (2010) High-throughput virtual screening using quantum mechanical probes: discovery of selective kinase inhibitors. ChemMedChem 5: 1007–1014.
17. Miyazaki Y, Nakano M, Sato H, Trucsdalc AT, Stuart JD, et al. (2007) Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro

evaluation targeting angiogenetic kinases. Bioorg Med Chem Lett 17: 250–254.
18. Bardelle C, Cross D, Davenport S, Kettle JG, Ko EJ, et al. (2008) Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. Bioorg Med Chem Lett 18: 2776–2780.
19. Lafleur K, Huang D, Zhou T, Caflisch A, Nevado C (2009) Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). J Med Chem 52: 6433–6446.
20. Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, et al. (2008) A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol 26: 127–132.
21. Fersht AR, Shi JP, Knill-Jones J, Lowe DM, Wilkinson AJ, et al. (1985) Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314: 235–238.
22. Hine J (1972) Structural Effects on Rates and Equilibria .15. Hydrogen-Bonded Intermediates and Stepwise Mechanisms for Proton-Exchange Reactions between Oxygen-Atoms in Hydroxylic Solvents. J Am Chem Soc 94: 5766–&.
23. Steiner T, Saenger W (1993) Role of C-H…O Hydrogen-Bonds in the Coordination of Water-Molecules - Analysis of Neutron-Diffraction Data. J Am Chem Soc 115: 4540–4547.
24. Cordero B, Gomez V, Platero-Prats AE, Reves M, Echeverria J, et al. (2008) Covalent radii revisited. Dalton Trans. pp 2832–2838.
25. Bondi A (1964) Van Der Waals Volumes+Radii. Journal of Physical Chemistry 68: 441–&.
26. Bohm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8: 243–256.
27. Morozov AV, Kortemme T, Tsemekhman K, Baker D (2004) Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. Proc Natl Acad Sci U S A 101: 6946–6951.
28. Momany FA, Rone R (1992) Validation of the General-Purpose Quanta(R)3.2/Charmm(R) Force-Field. J Comput Chem 13: 888–900.
29. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem 4: 187–217.
30. Warwicker J, Watson HC (1982) Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J Mol Biol 157: 671–679.
31. Im W, Beglov D, Roux B (1998) Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer Physics Communications 111: 59–75.
32. Gibson AE, Arris CE, Bentley J, Boyle FT, Curtin NJ, et al. (2002) Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O(6)-substituted guanine derivatives. J Med Chem 45: 3381–3393.
33. Stelmach JE, Liu L, Patel SB, Pivnichny JV, Scapin G, et al. (2003) Design and synthesis of potent, orally bioavailable dihydroquinazolinone inhibitors of p38 MAP kinase. Bioorg Med Chem Lett 13: 277–280.
34. Berger DM, Torres N, Dutia M, Powell D, Ciszewski G, et al. (2009) Non-hinge-binding pyrazolo[1,5-a]pyrimidines as potent B-Raf kinase inhibitors. Bioorg Med Chem Lett 19: 6519–6523.
35. Koch P, Jahns H, Schattel V, Goettert M, Laufer S (2010) Pyridinylquinoxalines and pyridinylpyridopyrazines as lead compounds for novel p38 alpha mitogen-activated protein kinase inhibitors. J Med Chem 53: 1128–1137.
36. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins 8: 195–202.
37. No KT, Grant JA, Scheraga HA (1990) Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method .1. Application to Neutral Molecules as Models for Polypeptides. Journal of Physical Chemistry 94: 4732–4739.
38. No KT, Grant JA, Jhon MS, Scheraga HA (1990) Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method .2. Application to Ionic and Aromatic-Molecules as Models for Polypeptides. Journal of Physical Chemistry 94: 4740–4746.
39. Irwin JJ, Shoichet BK (2005) ZINC–a free database of commercially available compounds for virtual screening. J Chem Inf Model 45: 177–182.
40. Hawkins PC, Warren GL, Skillman AG, Nicholls A (2008) How to do an evaluation: pitfalls and traps. J Comput Aided Mol Des 22: 179–190.
41. Yang CY, Sun H, Chen J, Nikolovska-Coleska Z, Wang S (2009) Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. J Am Chem Soc 131: 13709–13721.
42. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9: 430–431.

Supporting Information

Ranking the strength of individual hydrogen bonds in DNA base pairs

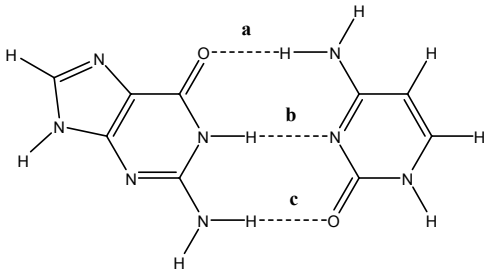
The three hydrogen bonds **a**, **b** and **c** in DNA bases pairs (Scheme S1) were ranked according to the hydrogen bonding strength calculated by $f_{hb}*(w_A + w_D)$. Different approaches [1,2,3,4,5,6] showed substantial discrepancy (Table S4). In the case of G-C complex, larger discrepancy was observed, presumably because the energies of the three hydrogen bonds are very close. However, in the case of A-T complex, five methods including us gave the same ordering, with **b** being the strongest, and all methods predicted **c** as the weakest.

Probing hydrogen bonds formed with bound water molecules

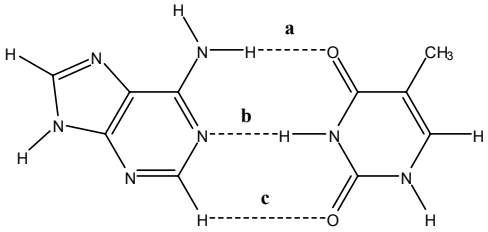
Among the interfacial water molecules observed in the HIV-protease-ligand complexes, Water 301 has been found to bridge the gaps between flaps of HIV-1 protease and inhibitors. This water molecule forms two hydrogen bonds with NH groups of Ile50 and Ile150 as an acceptor, meanwhile forms two additional hydrogen bonds with the carbonyl groups of the inhibitor as a donor (Fig. S7). Water 301 is observed in nearly all HIV-1 protease-ligand complexes, except in the case it is displaced deliberately [7].

To see whether the parameters employed can rigorously probe hydrogen bonds formed with implicit bound water molecules, four HIV-1 protease crystal structures were tested. As illustrated in Fig. S7, each of the four polar atoms forming hydrogen bonds with Water 301 is water accessible, and the calculated water molecules meet at the position of Water 301, indicative of one water molecule bridging all the four polar atoms. The number of predicted dots as potential water positions also represents the entropy of the water molecule. Compared with the other two crystal water molecules, the number of predicted dots for Water 301 is fewer, consistent with the fact that Water 301 is highly constrained. Hydrogen bonding penalty for each complex as well as partial contribution from the protein and ligand was given in Table S3. Only one of the catalytic aspartyl diad (Asp25 and Asp125) was deprotonated, as evidenced by NMR study [8].

Scheme S1. H-bonded complexes of nucleic acid pairs



a) Guanine-Cytosine (G-C)



b) Adenine-Thymine (A-T)

Scheme S2. Schematic view of probing hydrogen bonds formed with implicit water. Three shells around each donor or acceptor were generated with the radius being r_{sol}-0.15, r_{sol}, and r_{sol}+0.15 Å, respectively. In the current study, 2200 dots were distributed over the surface of each shell, with each dot representing the probability of a hydrogen bond with implicit water. Dots with the angle Θ less than 90° or within the inner shell of other atoms will be removed. Any dot left indicates hydrogen bonding with water remains.

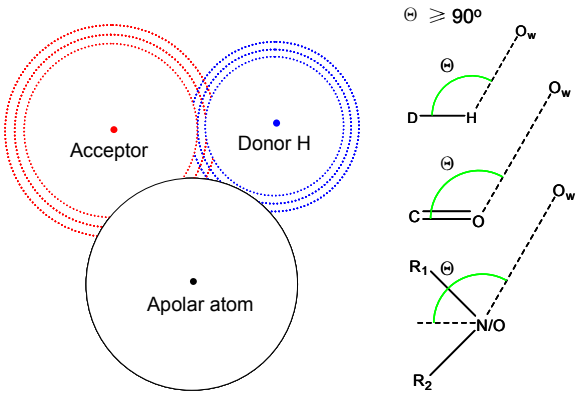


Table S1. MPEOE partial charge and water solubility of model small molecules used to generate initial guess of hydrogen bonding weights.

	MPEOE partial charge (e)	Solubility in water at 20 ^{o}C
pyridine	N (-0.58)	miscible
N-methylacetamide	O (-0.56); H (0.26)	N.A.
acetone	O (-0.49)	miscible
aniline	H (0.27)	3.6 g/100ml
n-propanol	O (-0.62); H (0.39)	miscible
n-butanol	O (-0.62); H (0.39)	7.7 g/100ml
phenol	O (-0.59); H (0.40)	8.3 g/100ml
diethyl ether	O (-0.37)	6.9 g/100ml

Table S2. Two-parameter LIECE energy and hydrogen bonding penalty on 24 EphB4 inhibitors [9].

No.	P_{HB} (kcal/mol)	LIECE (kcal/mol)	Exp. ΔG (kcal/mol)	No.	P_{HB} (kcal/mol)	LIECE (kcal/mol)	Exp. ΔG (kcal/mol)
8	1.05	-12.52	-7.93	**20**	1.04	-12.73	-7.79
9	1.03	-12.39	-7.23	**21**	1.09	-12.06	-6.40
10	1.07	-12.74	-8.10	**22**	1.05	-12.82	-8.54
11	1.02	-12.60	-7.69	**23**	1.01	-12.74	-8.05
12	1.09	-12.78	-6.75	**24**	1.49	-12.72	-6.07
13	1.04	-13.08	-8.02	**25**	1.05	-12.87	-10.09
14	1.02	-12.81	-7.12	**26**	1.03	-12.74	-8.90
15	2.09	-12.98	-6.62	**27**	1.43	-12.95	-5.99
16	1.04	-12.99	-8.65	**28**	1.02	-13.44	-9.61
17	1.04	-12.78	-7.91	**29**	1.05	-13.22	-8.46
18	1.31	-12.94	-7.38	**31**	1.03	-13.98	-8.51
19	1.03	-13.01	-8.45	**32**	1.06	-13.46	-8.19

Table S3. Hydrogen bonding penalty of HIV-1 protease complexes.

PDB code	Resolution (Å)	P_{HB}		
		Protein	Ligand	Total
1HXW	1.80	0.70	0.68	1.38
1HXB	2.30	0.58	0.66	1.23
1HPX	2.00	1.00	0.56	1.56
1HIH	2.20	1.04	0.11	1.16

Table S4. Ranking of the individual H-bonds strengths in the G-C and A-T base pairs (Scheme S1) on the basis of literature data and our results.

Method of characterizing the H-bonds	Calculation level	G-C	A-T
Rotation [1]	B3LYP/D95**	**c**>**b**>**a**	**a**>**b**>**c**
Compliance constants [2]	B3LYP/6-311++G**	**b**>**a**>**c**	**b**>**a**>**c**
Atom replacement [3]	B3LYP/6-311++G**	**a**>**c**>**b**	**b**>**a**>**c**
EML equation [4] for experimental geometry	B3LYP/6-311++G**	**a**>**b**>**c**	**a**>**b**>**c**
$E_{H\cdots B}$ vs. ρ_{CP} relation [5]	B3LYP/6-311++G**	**a**>**c**>**b**	**b**>**a**>**c**
NBO[6]	B3P86/6-311++G**	**a**>**b**>**c**	**b**>**a**>**c**
Hydrogen bonding weights	N.A.	**b**>**a**=**c**	**b**>**a**>**c**

Table S5. Hydrogen bonding penalty of the 100 kinase complex structures.

Protein	PDB code	P_{HB}	Protein	PDB code	P_{HB}	Protein	PDB code	P_{HB}
CDK2	1gz8	0.38	Abl	1iep	1.23	p38α	1ove	1.02
	1h0v	0.71		3kfa	1.57		1oz1	0.00
	1h0w	0.47		1opj	1.12		1w82	1.16
	3lfn	1.04		2f4j	1.83		1w84	0.21
	3lfq	0.24		2g2h	1.50		1yqj	0.00
	3lfs	0.71		2gqg	0.51		1ywr	1.11
	3ig7	0.06		3dk3	0.52		1zzl	1.27
	3igg	0.26		3dk6	0.52		3gc7	0.02
	1aq1	0.45		3cs9	0.22		3gcq	0.90
	1e1v	0.74		2qoh	0.61		3flq	1.01
	1e1x	0.71		2v7a	0.48		3fls	1.00
	1jvp	1.16		2z60	0.77		3flw	1.01
	1h1r	1.57		2e2b	1.47	AurA	3d15	1.11
	2r3f	0.49		2hyy	1.34		2c6e	1.76
	2r3h	0.11		2hz0	0.73		3d2k	1.12
	3le6	0.05		2hzi	1.50		3dj6	0.80
	1ke5	0.70	Lck	3bym	2.06		3e5a	1.00
EphB4	2vwu	1.51		3bys	0.21		3d14	1.20
	2vwx	1.18		2of2	0.83		3daj	0.00
	2vwy	1.96		2ofu	1.90		3bel	1.00
	2vwz	0.95		2ofv	0.54	EGFR	1xkk	0.20
	2vx0	0.90		2og8	0.74		2rgp	1.58
	2vx1	0.98		1qpj	0.99	Tie2	2oo8	0.53
	2vwv	1.18	JAK2	3e62	0.13		2p4i	0.96
	2vww	0.96		3e63	0.02	PAK4	2x4z	1.02
JNK3	1pmn	1.92		3e64	0.52		2cdz	0.10
	1pmv	0.30		3iok	1.19	Kit	1t46	1.70
	2b1p	0.16		3kck	0.56		2oiq	0.12
	2ok1	0.27		3jy9	0.29	BTK	3gen	0.81
	2p33	1.29		3lpb	2.01		3k54	1.02
	2r9s	0.32	Ret	2x2l	0.00	Hck	1qcf	0.61
	3g90	2.04		2ivu	1.08		2hk5	0.28
	3fi3	0.39		2ivv	0.41			
	3g9l	1.46		2x2m	0.11			

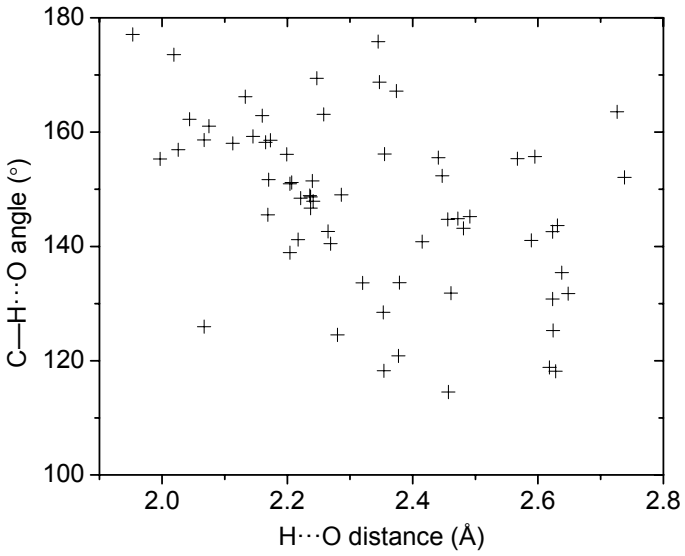


Figure S1. Scatter plot of C—H···O angles against H···O distances in short C—H···O interactions between ligands and proteins.

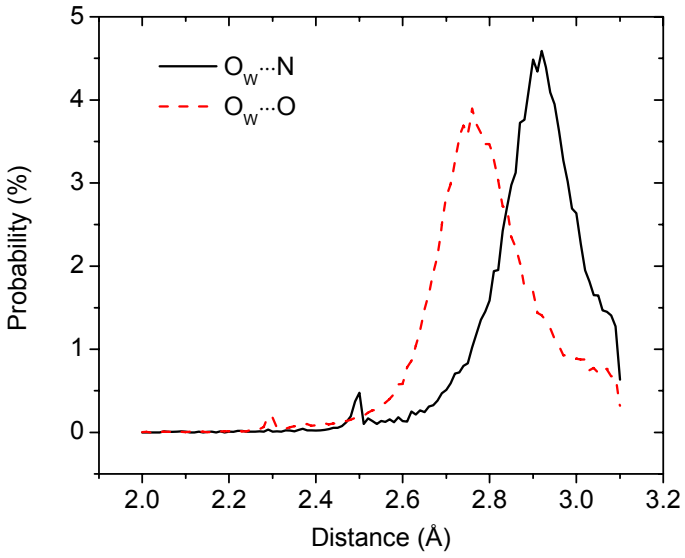


Figure S2. Distribution of distances between crystal water oxygen and oxygen or nitrogen atoms of proteins.

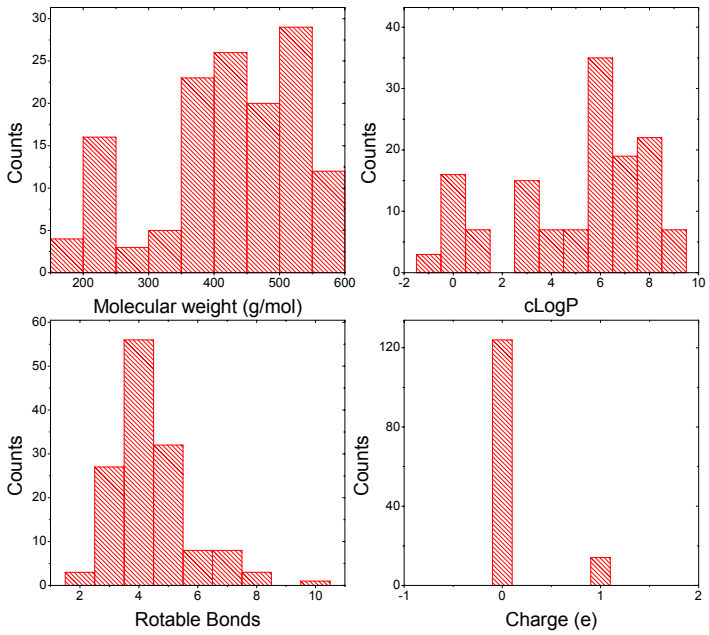


Figure S3. Distribution of some key properties of the inhibitors used in the training and test set.

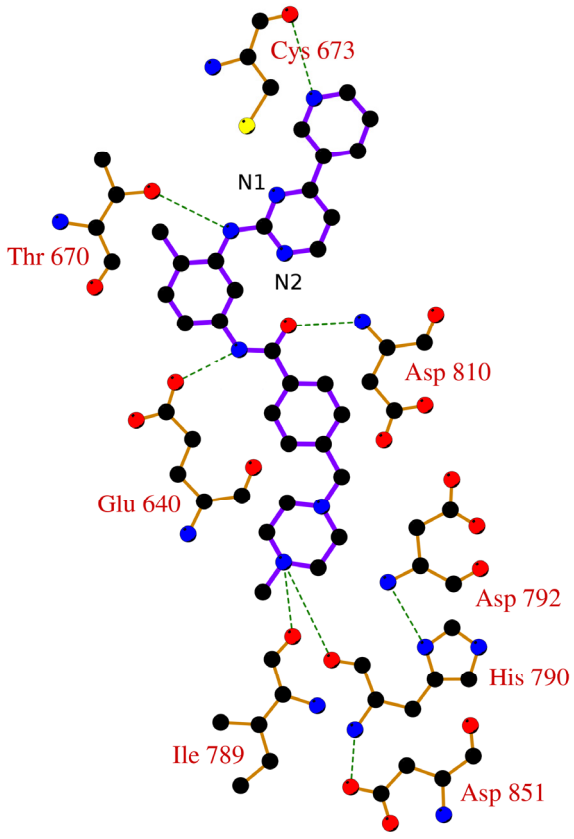


Figure S4. 2D plot of the binding mode of Imatinib. Upon ligand binding, one nitrogen atom of the Imatinib pyrimidine ring (N1) becomes water inaccessible and does not form a new hydrogen bond, leading to a penalty of 1. By contrast, the other nitrogen atom (N2) remains hydrogen bonding with a nearby bound water molecule and thus has no penalty.

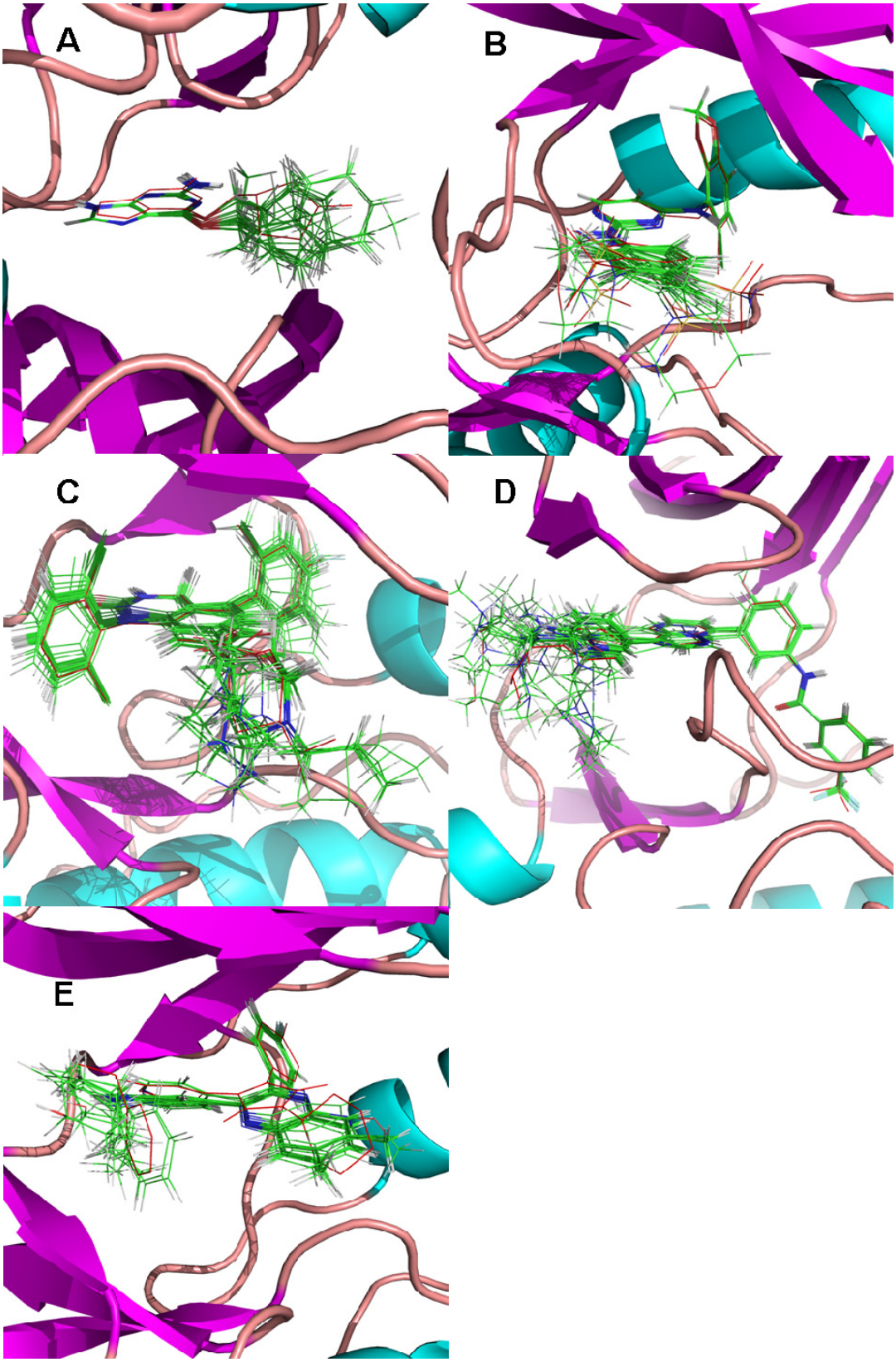


Figure S5. Poses with the most favorable binding energy of inhibitors of CDK2 (A), EphB4 (B), p38 α (C), Braf (D) and another set of p38α inhibitors (E). The molecules with bonds in red are the binding modes of the corresponding scaffolds in the crystal structures.

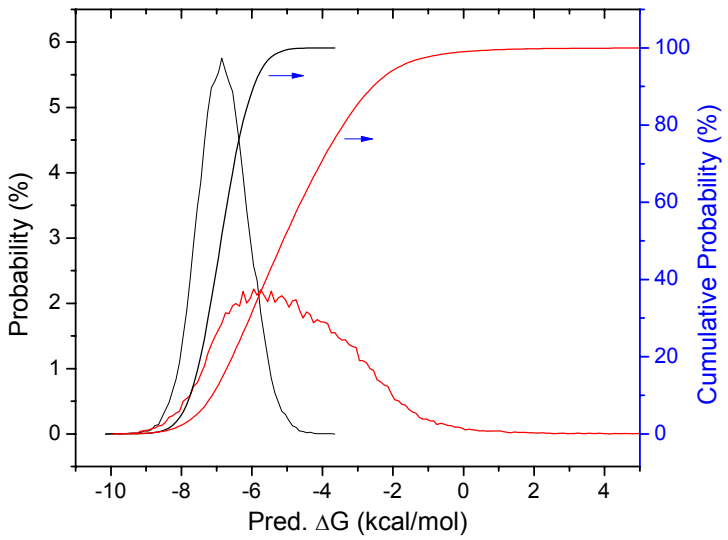


Figure S6. Distribution of predicted binding affinities by Autodock4 (black) and the proposed scoring function (red) on 74,678 compounds passing the first two filters (HB to Met696 and $P_{HB} \leq 2$ kcal/mol). Bin size: 0.1 kcal/mol.

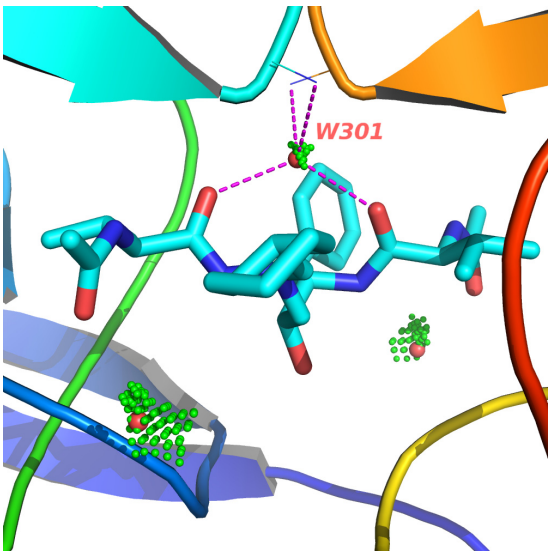


Figure S7. Cartoon representation of HIV-1 protease ligand site (1HIH) with crystal (red dots) and calculated (green) water oxygen atoms.

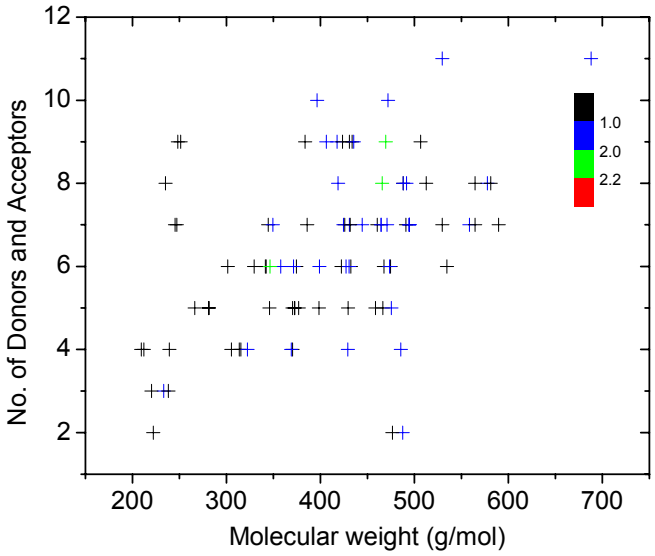


Figure S8. Number of donors/acceptors against molecular weight of inhibitors complexed in the 100 kinase crystal structures. Colored according to hydrogen bonding penalty.

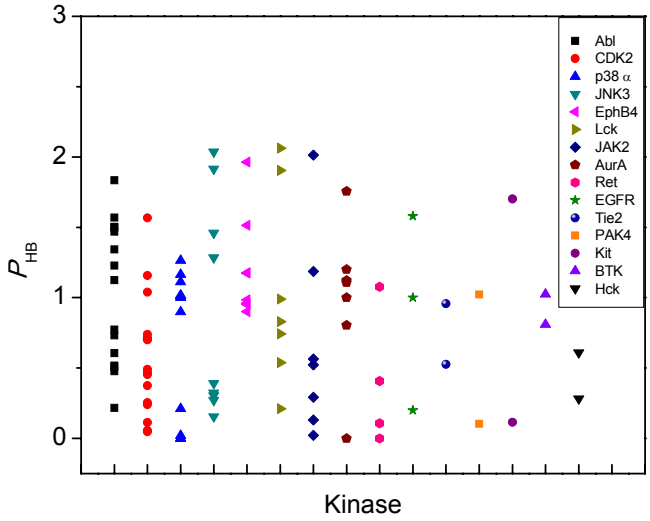


Figure S9. Statistics of hydrogen bonding penalty on the 100 kinase complexes.

References

1. Asensio A, Kobko N, Dannenberg JJ (2003) Cooperative hydrogen-bonding in adenine-thymine and guanine-cytosine base pairs. Density functional theory and Moller-Plesset molecular orbital study. Journal of Physical Chemistry A 107: 6441-6443.
2. Grunenberg J (2004) Direct assessment of interresidue forces in Watson-Crick base pairs using theoretical compliance constants. J Am Chem Soc 126: 16310-16311.
3. Dong H, Hua W, Li S (2007) Estimation on the individual hydrogen-bond strength in molecules with multiple hydrogen bonds. J Phys Chem A 111: 2941-2945.
4. Matta CF, Castillo N, Boyd RJ (2006) Extended weak bonding interactions in DNA: pi-stacking (base-base), base-backbone, and backbone-backbone interactions. J Phys Chem B 110: 563-578.
5. Ebrahimi A, Khorassani SMH, Delarami H (2009) Estimation of individual binding energies in some dimers involving multiple hydrogen bonds using topological properties of electron charge density. Chemical Physics 365: 18-23.
6. Szatylowicz H, Sadlej-Sosnowska N (2010) Characterizing the Strength of Individual Hydrogen Bonds in DNA Base Pairs. J Chem Inf Model.
7. Lam PY, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, et al. (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 263: 380-384.
8. Wang YX, Freedberg DI, Yamazaki T, Wingfield PT, Stahl SJ, et al. (1996) Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. Biochemistry 35: 9945-9950.
9. Bardelle C, Cross D, Davenport S, Kettle JG, Ko EJ, et al. (2008) Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. Bioorg Med Chem Lett 18: 2776-2780.

Chapter 3

Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics

CHEMMEDCHEM FULL PAPERS

DOI: 10.1002/cmdc.201200331

Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics

Hongtao Zhao, Danzhi Huang,* and Amedeo Caflisch*[a]

Several small molecules that bind to the inactive DFG-out conformation of tyrosine kinases (called type II inhibitors) have shown a good selectivity profile over other kinase targets. To obtain a set of DFG-out structures, we performed an explicit solvent molecular dynamics (MD) simulation of the complex of the catalytic domain of a tyrosine kinase receptor, ephrin type-A receptor 3 (EphA3), and a manually docked type II inhibitor. Automatic docking of four previously reported type II inhibitors was used to select a single snapshot from the MD trajectory for virtual screening. High-throughput docking of a pharmacophore-tailored library of 175 000 molecules resulted in about 4 million poses, which were further filtered by van der Waals efficiency and ranked according to a force-field-based energy function. Notably, around 20% of the compounds with predicted binding energy smaller than -10 kcal mol^{-1} are known type II inhibitors. Moreover, a series of 5-(piperazine-1-yl)isoquinoline derivatives was identified as a novel class of low-micromolar inhibitors of EphA3 and unphosphorylated Abelson tyrosine kinase (Abl1). The in silico predicted binding mode of the new inhibitors suggested a similar affinity to the gatekeeper mutant T315I of Abl1, which was verified in vitro by using a competition binding assay. Additional evidence for the type II binding mode was obtained by two 300 ns MD simulations of the complex between *N*-(3-chloro-4-(difluoromethoxy)-phenyl)-2-(4-(8-nitroisoquinolin-5-yl)piperazin-1-yl)acetamide and EphA3.

Introduction

Protein kinases represent attractive targets in oncology drug discovery.[1] One such target is Abelson tyrosine kinase (Abl1), for which small-molecule drugs are employed in the clinics to treat chronic myelogenous leukemia (CML). However all current drugs including imatinib, nilotinib and dasatinib are incapable of inhibiting the most notable T315I gatekeeper mutant, detected in 10–20% of patients with CML after failure of imatinib therapy.[2,3] Another interesting class of targets is the erythropoietin-producing human hepatocellular carcinoma receptors (Eph), the largest family of receptor tyrosine kinases. The Eph receptors have been implicated in sprouting angiogenesis and blood vessel remodeling during vascular development.[4–7] Furthermore, overexpression of several of the 14 known Eph receptors, including ephrin type-A receptor 3 (EphA3), has been linked to tumors and the associated vasculature, suggesting a critical role in tumor-related angiogenesis.

The majority of small-molecule kinase inhibitors developed so far target the ATP binding site of the kinase in its active state (DFG-in), and are known as type I inhibitors.[8] However, the first kinase-targeting small molecule to reach the market was imatinib (Gleevec), a type II tyrosine kinase inhibitor that binds to the inactive state of Abl1 characterized by a closed conformation of the activation loop (DFG-out). The flip of the DFG motif, a conserved triad (Asp–Phe–Gly) at the beginning of the activation loop, induces remarkable changes in the ATP binding site and exposes an additional hydrophobic pocket that is less conserved in sequence.[9] Many kinase inhibitors have failed in preclinical or clinical development due to their lack of selectivity causing intolerable side effects, largely because the kinase ATP binding site is highly conserved in sequence and conformation.[10] The emergence of type II inhibitors creates new opportunities by targeting the allosteric pocket of the DFG-out conformation, offering selectivity and intellectual property novelty.[11]

Structure-based virtual screening of type II inhibitors requires experimentally available DFG-out protein structures, which were initially limited in availability. As a result, most known type II inhibitors to date have been developed via quantitative structure–activity relationship (QSAR)-guided modifications of ATP binding site ligands.[8] Several computational approaches have been proposed to convert a kinase from a DFG-in into a DFG-out conformation, such as DOLPHIN by deleting about six residues of the activation loop starting with the DFG motif.[12] More recently, a protein remodeling program has been used to model a DFG-out conformation by using the DFG-in as a template structure.[13] At present, a few pharmaceutically relevant kinases have been co-crystallized with type II inhibitors. Even so, literature reports describing the discovery of type II inhibitors by virtual screening remain rare. One reason for this is an induced fit in the protein X-ray structure in favor of the co-crystalized inhibitor; if the biased bind-

[a] H. Zhao, Dr. D. Huang, Prof. A. Caflisch
Department of Biochemistry, University of Zurich
Winterthurerstrasse 190, 8057 Zurich (Switzerland)
E-mail: dhuang@bioc.uzh.ch
caflisch@bioc.uzh.ch

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201200331.

CHEMMEDCHEM

ing site of a single X-ray structure is used, poor docking results are common when using structurally diversified ligands. The unfavorable interactions, particularly clashes between putative ligands and the kinase, on the other hand, would be accommodated by the rearrangement of the protein target, due to the plasticity of the loops around the ATP binding site. The failure of docking diversified ligands thus calls for a generalized binding site in structure-based virtual screening, in order to explore a larger chemical space. The discovery of type II inhibitors is further complicated in that such inhibitors target a relatively scarcely populated protein conformation,[14] which is presumably kinase dependent.[15]

Here, we report the identification of type II inhibitors by flexible ligand docking into an EphA3 structure generated by molecular dynamics (MD). We first run a constrained MD simulation with explicit solvent to induce a fit of the EphA3 structure to a known type II inhibitor that could not be docked into the original X-ray structure. In silico screening was then carried out by pharmacophore filtering, high-throughput docking, and ranking based on an energy function with continuum solvation and hydrogen bonding penalty. Retrospectively, we identify ten classes of known type II scaffolds, none of which could be discovered based on the original X-ray structure and among which some are reported to be active against EphA3. Prospectively, our endeavors lead to the identification of a novel class of low micromolar type II inhibitors, which retains inhibitory activity against the T315I gatekeeper mutant of Abl1.

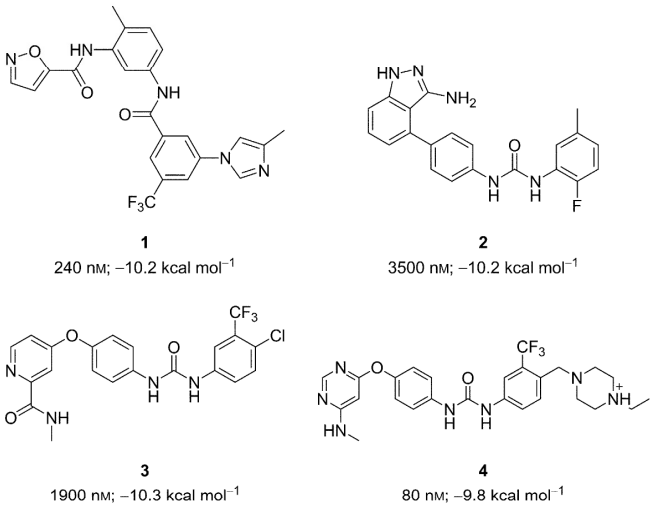


Figure 1. Previously reported type II inhibitors of EphA3.[16,17] The values next to the compound number are the experimentally measured dissociation constant against phosphorylated EphA3 and the predicted binding free energy. The latter was calculated using the MD-IF structure and a scoring function with continuum solvation and hydrogen bonding penalty.[21]

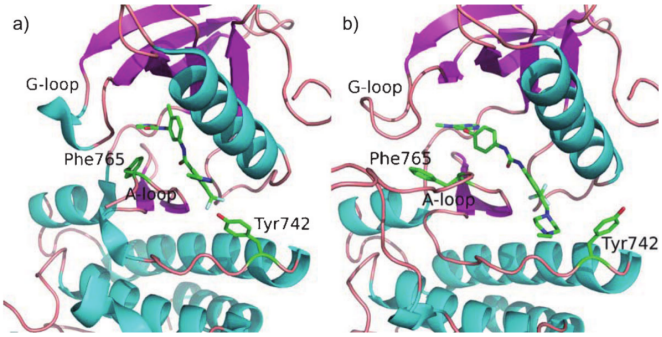


Figure 2. Comparison of a) the crystal structure (PDB: 3DZQ) of the complex of EphA3 with inhibitor **1** and b) the binding mode obtained by docking compound **4** into the MD-IF structure. The comparison shows the different orientations of Tyr742 and Phe765, and the difference in the G-loop.

Results and Discussion

Inducing a generalized DFG-out conformation by MD

Induced fit in favor of a specific inhibitor exists in the X-ray crystal structure of EphA3 co-crystallized with compound **1** (PDB: 3DZQ),[16] as none of compounds **2** to **4**[17] (Figure 1) can be docked into the X-ray structure in a type II binding mode without clashes. Specifically, the glycine-rich loop (G-loop), which can adopt various conformations,[10] collapses into the ATP binding site and tightly encompasses the small type I head group of compound **1**. As a consequence, the bigger type I head groups of compounds **2** to **4** in the ATP binding site would clash with the G-loop. In addition, the side chain of Tyr742 blocks the entry of the piperazine group of compound **4** (Figure 2a). Experimentally, different orientations of the Tyr742 side chain have been reported in the X-ray structures of DFG-in EphA3 (PDB: 2QOB[18]) and EphA4 (PDB: 2Y6O and 2Y6M[19]). Computationally, we have observed that the Tyr742 side chain can adopt two distinct, equally populated orientations in the DFG-out conformation based on ten 50 ns explicit solvent MD simulations of EphA3 (PDB: 3DZQ) with a trifluoromethylbenzene in the allosteric site (Figure S1 in the Supporting Information). These MD simulations were carried out using the protocols described in our previous work,[20] and a detailed analysis of these results will be presented elsewhere.

In the present work, the χ_1 angle of the side chain of Tyr742 was rotated from -173° (as in 3DZQ) to -60°, and compound **1** was manually replaced by compound **2** to obtain a generalized binding site that can accommodate diversified ligands. Explicit solvent MD with harmonic constraints on all C_α atoms ex-

cluding the G- and A-loops was then carried out for 2 ns. The first 1 ns segment of the trajectory was discarded, and seven snapshots in the second half of the trajectory were selected as they preserve the five intermolecular hydrogen bonds shown in Figure 3. These seven snapshots were minimized over 200

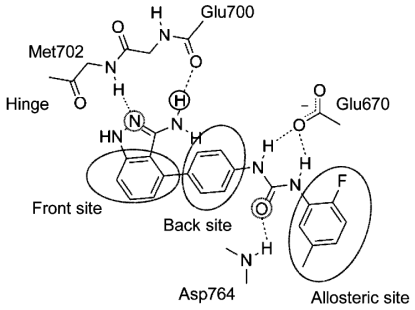


Figure 3. Pharmacophore mapping of the key interactions of type II kinase inhibitors illustrated by compound **2** and EphA3. Hydrogen bonds are shown as dashed lines. Pharmacophore elements used to filter the ZINC library: two acceptors (dashed circles), one donor (solid circle), and three hydrophobic rings (ovals). The hydrogen bond to Glu 670 was not used as a pharmacophore because of the flexibility and solvent exposure of the Glu 670 side chain (see text for discussion). Details of the geometric constraints are illustrated in Figure S2 in the Supporting Information.

steps, and then compound **2** together with all water molecules were removed. Flexible docking of compounds **1**–**4** into each of the seven structures was further used to select a single structure for screening according to the binding affinity, which was estimated by a previously reported scoring function using exactly the same parameters.[21] The selected snapshot is called the molecular dynamics induced fit (MD-IF) structure. It should be pointed out that the induced displacements of the G- and A-loops and reorientation of Tyr 742 are not achievable by simple energy minimization.

Pharmacophore tailoring the ZINC library

The majority of kinase inhibitors—including type II inhibitors—are hinge binders. They usually form a key hydrogen bond with the backbone NH of the hinge, which belongs to Met 702 in the case of EphA3. A hydrogen bond with the carbonyl oxygen of Glu 700 is also observed, including the acidic CH groups as donors.[21] For type II inhibitors, an additional pair of hydrogen bonds can be observed with Asp 764 of the DFG motif, and the catalytically important Glu 670 from the αC-helix (Figure 3).[8] However, the hydrogen bond with Glu 670 is surrounded by water molecules, and this hydrogen bond is kinetically not stable.[22] Aside from the hydrogen-bonding interactions, most type II inhibitors can be mapped well into three major hydrophobic interactions: ATP front site, ATP back site, and the allosteric site (Figure 3). The combination of three hydrogen bonds with three hydrophobic groups and their relative separations used as constraints (Figure S2 in the Supporting Information) reduces the 9 millions compounds in the ZINC library (August 2011) to less than 200 000 molecules (Figure 4) within 6 h on a single Xeon 2.8 GHz central processing unit (CPU).

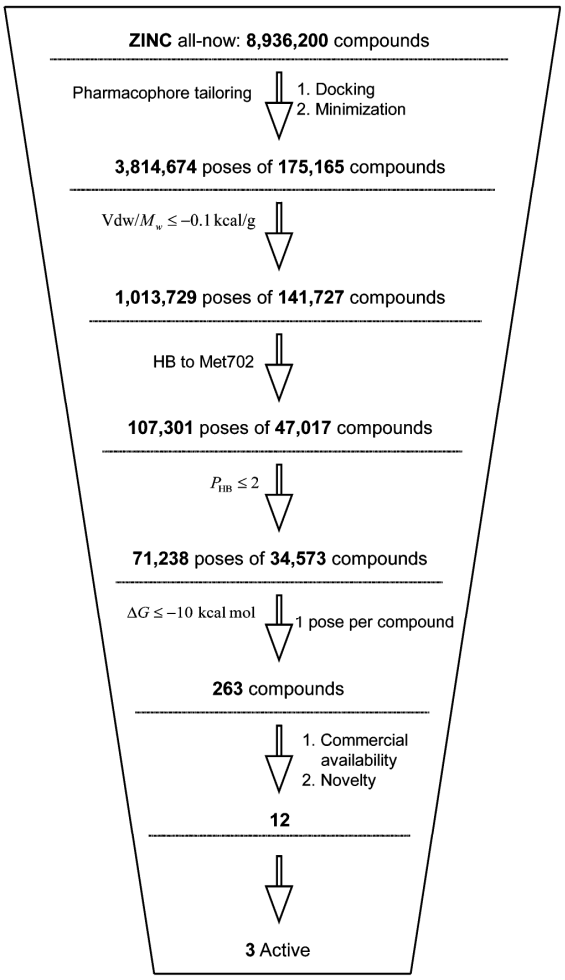


Figure 4. Schematic illustration of the high-throughput virtual screening process. The pharmacophore tailoring required 6 h, while the docking, minimization and evaluation of ΔG of binding were carried out on a computer cluster of 300 cores within one week, which corresponds to approximately 1500 days, 300 days and 150 days, respectively, of the equivalent of a single commodity processor.

Structure-based flexible ligand docking

The docking of 175 165 compounds by AutoDock yielded about 3.8 millions poses. To improve computational efficiency, three filters were applied to these poses, with focus on three complementary aspects: potency, binding specificity, and hydrogen-bonding conditions of polar atoms (Figure 4). The van der Waals efficiency of -0.1 kcal g^{-1}[23] was used as the first filter. Secondly, as the majority of kinase inhibitors are hinge binders, the hydrogen bond with the NH group of Met 702 was used as the second filter to gain binding specificity, which

CHEMMEDCHEM

is most efficient among the applied filters. Lastly, a hydrogen-bonding penalty[21] of two was used to remove poses that have polar atoms buried in hydrophobic sites. With these three filters, the number of poses was reduced to about 71 000 of 35 000 compounds. Finally, the previously reported scoring function[21] was applied with a cutoff value of -10 kcal mol^{-1}, which yielded 263 compounds for further evaluation (Figure 4).

Evaluation of screening results

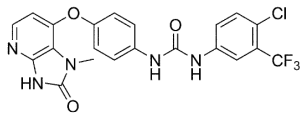
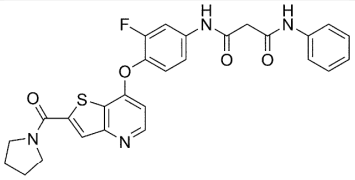
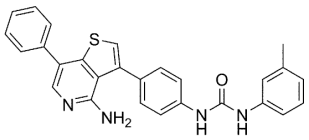
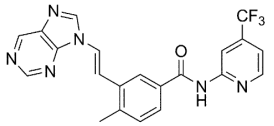
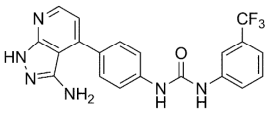
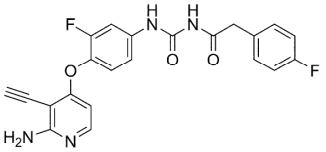
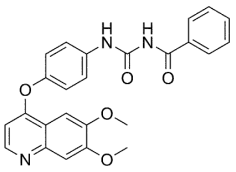
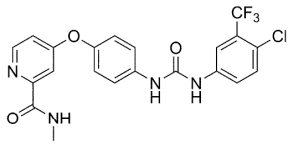
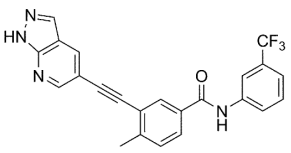
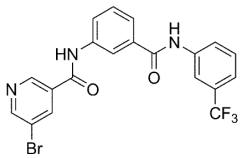
Interestingly, among the top 263 compounds, 55 (21%) are known type II inhibitors of 10 different scaffolds, primarily targeting Braf, Met, VEGFR, Abl1, SRC, Tie-2 and Eph (Table 1).[16,24–34] Kinase inhibitors, including type II, typically exhibit cross activity on a subfamily, as observed experimentally.[35] Indeed, some of the known inhibitors in the top 263 compounds are reported to show low micromolar to nanomolar activity on EphA3.[16,17] The successful recovery of structurally diversified known type II scaffolds out of millions of compounds indicates that the use of an MD-IF structure is very effective. Notably, none of the above compounds can be docked as type II into the original X-ray structure without clashes with the G-loop, providing further evidence for the usefulness of inducing a generalized binding site by MD-based sampling.

In vitro validation

Twelve compounds were further selected for experimental validation, based on their novelty and commercial availability. Binding affinity was measured by a phage-display-based competition assay[17] (see Experimental Section). Three of the twelve compounds (piperazine derivatives **5**, **6** and **8**; Table 2) showed micromolar activity against phosphorylated EphA3. Unfortunately, it is not possible to carry out the same assay on unphosphorylated EphA3 due to a lack of commercial availability. It has been suggested that differential binding to phosphorylated and unphosphorylated forms of Abl1 can functionally differentiate compounds that prefer an inactive DFG-out kinase conformation (type II inhibitors) from those that do not (type I inhibitors), even for compounds that are not primarily Abl1 inhibitors but exhibit modest affinity for Abl1.[36] Interestingly, piperazine derivatives **5**, **6**, and **8** are active on unphosphorylated Abl1 (Table 2) and inactive on phosphorylated Abl1 (Table S1), which provides further evidence that they are type II inhibitors.

The nitro group of these compounds is not predicted to be involved in binding (Figure 5a) and so could be neglected for hit optimization. Compound **5** was used as the query scaffold for a similarity search that yielded a set of 20 derivatives in the Enamine library. Four of these 20 derivatives (compounds **7** and **9**–**11**) show micromolar affinity for unphosphorylated Abl1 (Table 2).

Table 1. Representative structures of previously reported type II inhibitors identified by virtual screening of the ZINC library using the MD-IF structure of EphA3 for docking.

Core structure	Primary target	Core structure	Primary target
	Braf[24–26]		c-Met VEGFR2[27]
	VEGFR2[28]		Abl1 SRC[29]
	VEGFR PDGFR[30]		Met[31]
	c-Met[32]		Braf VEGFR[33]
	Tie-2[34]		Similar to type II inhibitors of EphA3[16]

Binding to the T315I Abl1 mutant

In the predicted binding mode obtained by docking and validated by MD (see below), the piperazine group of compound **5** is away from the gatekeeper Thr699 (Figure 5a). Automatic structural alignment of the EphA3 MD-IF structure to the complex of Abl1 with the type II inhibitor DCC-2036 (PDB: 3QRJ[2]) shows that the hydrogen bonding donors and accept-

Table 2. Affinities (%) of compounds **5**–**11** for wild-type and mutant tyrosine kinases measured by phage-display-based competition binding assay.[a]

Compd	Substituents 3	Substituents 4	EphA3 wild type	Abl1-unphosph wild type	Abl1-unphosph T315I
5	Cl	$OCHF_2$	53	12 (8.5 μM)[c]	45
6	Br	H	63	48	>65
7	Cl	CH_3	61	48	>65
8[b]	CF_3	H	>65	2.8 (3.9 μM)[c]	5.8
9[b]	Cl	OCH_3	>65	28	>65
10[b]	Cl	F	>65	38	>65
11[b]	Cl	H	60	21	55

[a] Percentage of kinase not displaced by the test compounds at 30 μM concentration with small values indicating high affinity.[17] Compounds **5**, **6**, and **8** were identified by high-throughput flexible ligand docking into the MD-IF conformation of the EphA3 receptor tyrosine kinase. Compounds **7** and **9**–**11** were identified by similarity search using the scaffold of compound **5**. [b] The indicated compounds are racemic mixtures, purity was checked by in-house ESI-MS and then compounds were used as purchased. [c] The K_d value in parentheses is the mean of two dose–response measurements of 11 points each.

ors, as well as the rings of compound **5**, are positioned and oriented in a similar way (Figure 5b). The piperazine group of compound **5** is a little further away from the gatekeeper residue Ile315 than the phenyl ring of DCC-2036 in structure 3QRJ. Since the affinity of DCC-2036 is only slightly affected by the notable T315I gatekeeper mutant of Abl1,[2] we speculate that compound **5** might also bind to this mutant, which is the predominant mechanism of drug-induced resistance in imatinib-treated patients. Indeed, in the competition binding assay,[17] compound **8** at 30 μM shows a percent control (%ctrl) value of 5.8% on unphosphorylated T315I Abl1, which is very close to the value of 2.8% on unphosphorylated wild-type Abl1 (Table 2).

Validation of binding mode by MD simulations

To provide further evidence of the binding mode of compound **5**, two explicit solvent MD simulations were performed, starting from the complex with the MD-IF structure as obtained by docking (Figure 6; see also Figure S3 in the Supporting Information). Moreover, two MD runs were started from the X-ray structure of the complex between compound **1** and EphA3 (PDB: 3DZQ) as a basis for comparison. Overall, the binding modes of both compounds are stable

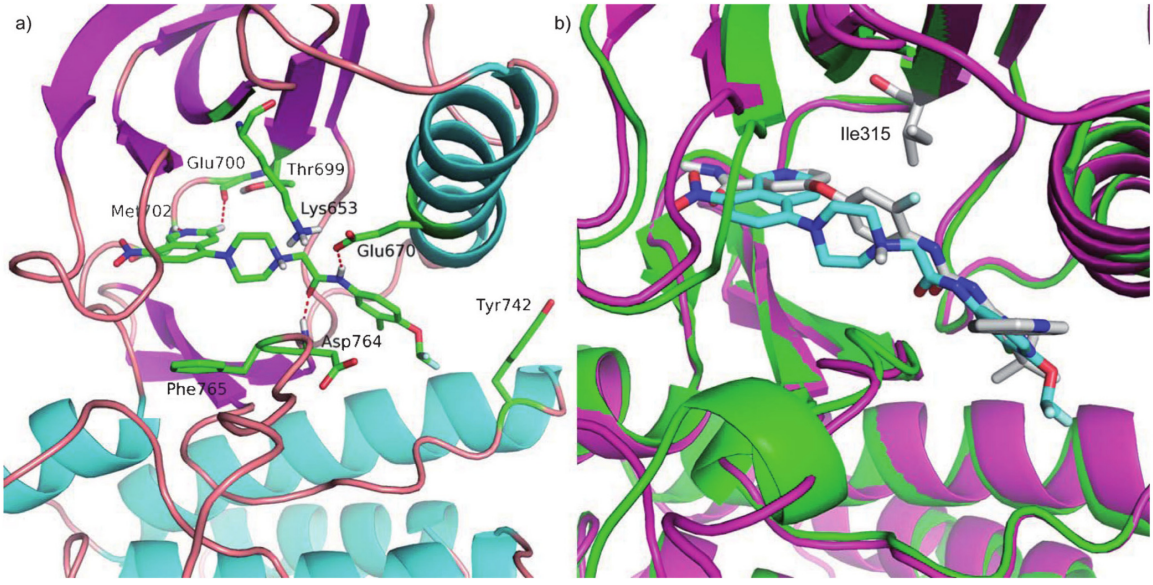


Figure 5. Predicted binding mode of compound **5** in the MD-IF structure of a) EphA3 and b) in the Abl1 crystal structure. a) Compound **5** and the EphA3 side chains involved in binding are shown by atom type coloring with carbon atoms in green, nitrogen atoms in blue, and oxygen atoms in red. b) The binding mode of compound **5** (carbon atoms in cyan) in Abl1 is obtained by structural superposition of the C_α atoms of the MD-IF EphA3 structure (magenta) into the crystal structure (PDB: 3QRJ) of Abl1 (green) in complex with the "switch control" inhibitor DCC-2036[2] (carbon atoms in gray). The superposition suggests that a decrease in affinity for the T315I mutant of Abl1 would not occur because there are no contacts between compound **5** and the Ile315 side chain.

CHEMMEDCHEM

(Figure S5 in the Supporting Information). The hydrogen bond between compound **1** and the hinge is broken in one of the two MD runs, which is probably due to the electrostatic repulsion between the isoxazole oxygen and the carbonyl oxygen of Glu700. In contrast, the two polar interactions of compound **5** with the hinge, the hydrogen bond to the NH of Met702 and CH···O=C bond involving the carbonyl of Glu700 (d1 and d2 in Figure 6; see also Figure S3 in the Supporting Information), are stable in both 300 ns MD simulations except for some transient ruptures. Both compounds are predicted to form stable hydrogen bonds with the DFG motif (d3 and d4; Figure 6), while the hydrogen bond between the amide NH of compound **1** or **5** and the carboxylic group of Glu670 (d5) fluctuates strongly because the latter is solvent exposed. For this reason, the hydrogen bond to Glu670 was not used as a pharmacophore to prefilter the ZINC library, although it is frequently observed with type II inhibitors. Nevertheless, the Glu670 fluctuations are smaller for compound **5** than **1**, which is probably due to the favorable electrostatic interactions with the positively charged piperazine ring of compound **5**.

Conclusions

To obtain an inactive DFG-out conformation for high-throughput docking, we have run an explicit solvent MD simulation of the EphA3 receptor tyrosine kinase with a type II inhibitor placed manually in the ATP binding site upon removal of the original inhibitor of the crystal structure. Along the MD trajectory, a snapshot that accommodates four previously reported type II inhibitors was selected for high-throughput flexible ligand docking. The docked library consisted of about 175000 compounds derived from nearly 9 million molecules using two-dimensional chemical descriptors and three-dimensional geometric constraints (i.e., relative distance and orientation of

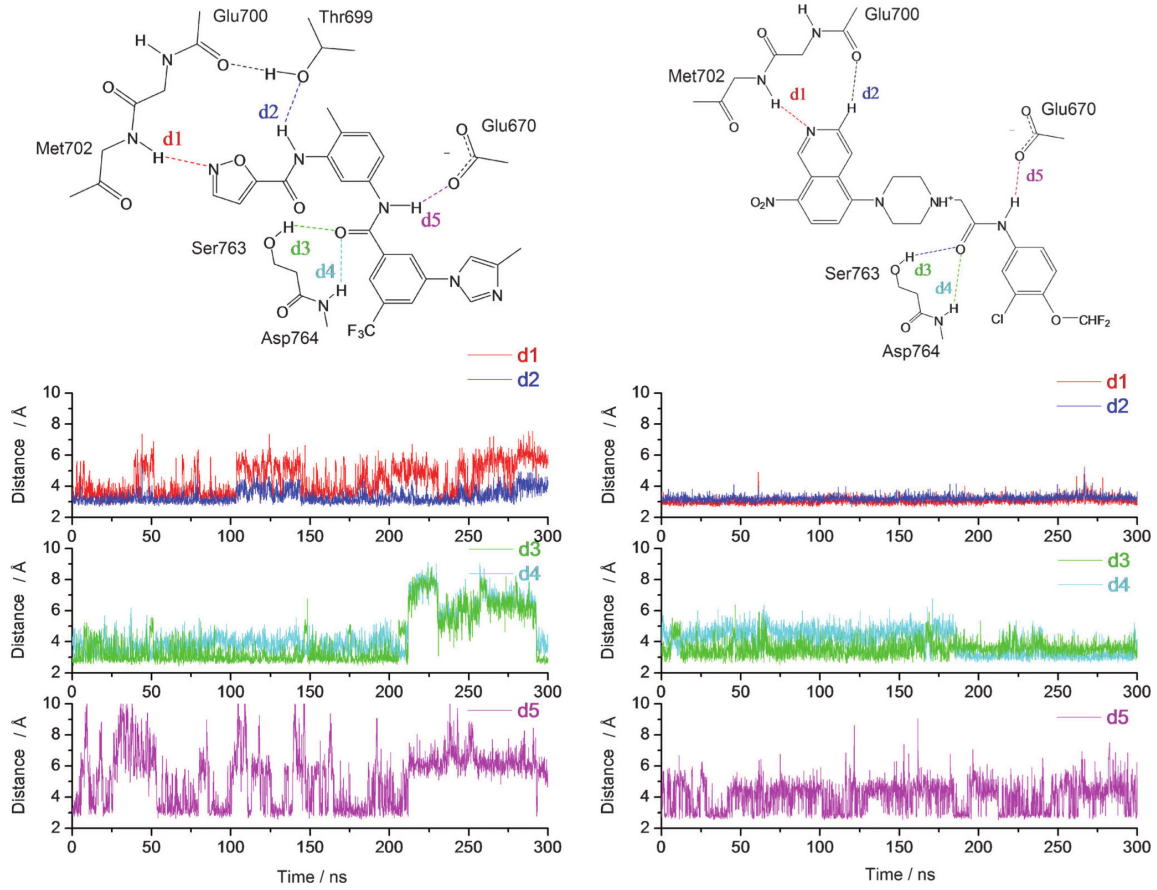


Figure 6. Explicit solvent MD validation of binding mode of compound **5**. (Left) MD run started from the X-ray structure of the complex between EphA3 and inhibitor **1** (PDB: 3DZQ). (Right) MD run started from the docked pose of compound **5** into the MD-IF structure of EphA3. (Top) Two-dimensional illustrations; (Bottom) time evolutions of hydrogen bonding distances measured between donor (or isoquinoline C_3 atom of compound **5**) and acceptor atoms. In the simulation with inhibitor **1**, the amide group associated with d3, d4 and d5 flipped by 180° at about 230 ns (see Figure S4 in the Supporting Information) and flipped back at 290 ns. The d5 distance shows large oscillation because of fluctuation of the Glu670 side chain. Note also that values around 5 Å reflect single water bridged hydrogen bonds with Glu670. Another MD run for each of compounds **1** and **5** is shown in Figure S3 in the Supporting Information. Detailed analysis is shown in Figures S4 and S5 in the Supporting Information.

pairs of functional groups). Using this procedure, we have identified a series of 5-(piperazine-1-yl)isoquinoline derivatives that exhibited low micromolar affinities for unphosphorylated Abl1 in a competition binding assay.

The following experimental evidence and computational results support the binding mode, which is predicted to be of type II: 1) The discovered scaffold shows a higher affinity for inactive than active Abl1 tyrosine kinase, a typical feature of type II inhibitors;[36] 2) the docking pose suggests a similar affinity for the gatekeeper mutant T315I as for wild-type Abl1, which was verified experimentally; 3) the hydrogen-bonding pattern and overall binding mode are preserved in two 300 ns MD simulations, started from the pose obtained by docking.

In conclusion, we have discovered a novel chemical class of type II tyrosine kinase inhibitors by using an in silico procedure based on a combination of explicit water MD simulations and high-throughput docking.

Experimental Section

Computational methods

MD simulations: The coordinates of missing atoms in the EphA3 crystal structure (PDB: 3DZQ), especially the long activation loop, were generated by the program Modeller (version 9.0).[37] To reproduce physiological pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysine and arginine residues were positively charged, while all other residues were considered neutral. The MD simulations were performed with the program NAMD (version 2.7)[38] using the all-atom CHARMM PARAM27 force field,[39] the TIP3P model of water,[40] and the CHARMM general force field for small molecules.[41]

The protein–ligand complexes were inserted into a cubic water box, with a minimal distance of 12 Å between any solute atom and the boundary of the box. Chloride and sodium ions were added to neutralize the system and to give an approximate salt concentration of 150 mM. If the distance between the water oxygen and any atom of the complex or any ion was smaller than 2.4 Å, the water molecules overlapping with the solute atoms or the ions were removed. Periodic boundary conditions were applied to avoid finite-size effects. Electrostatic interactions were calculated within a cutoff of 10 Å, while long-range electrostatic effects were taken into account by the particle mesh Ewald summation method.[42] Van der Waals interactions were treated with the use of a switch function starting at 8 Å and turning off at 10 Å. The temperature was kept constant at 310 K by using the Langevin temperature control with a damping coefficient of 1 ps^{-1}, while the pressure was held constant at 1 atm by applying a pressure piston. Before the production runs, water molecules and ions were subjected to energy minimization for 6000 steps, and a 1 ns equilibration with harmonic constraints (1 kcal mol^{-1} $Å^{-2}$) applied to the positions of protein C_α atoms excluding the G- and A-loop (residues 628–634 and 768–788, respectively). Covalent bonds involving hydrogen atoms were constrained by means of the SHAKE algorithm, and the dynamics were integrated with a time step of 2 fs.

Compound library: The compounds were downloaded from ZINC all-now library.[43] The library was firstly filtered by physicochemical properties, such as, number of donors and acceptors, molecular weight, and number of rings, according to the pharmacophores defined in Figure 3. Preparation included the assignment of CHARMm atom types, force field parameters,[44] and partial charges,[45] and energy minimization with a distance dependent dielectric function using the program CHARMM.[46,47] Finally, the pharmacophore software LIBO version 1.0 (Zhao and Calfisch, unpublished) developed in-house was used to filter the library by pharmacophore constraints according to Figure 3.

Docking: AutoDock (version 4.0)[48] was used to generate the binding poses over the conformational search space using the Lamarckian genetic algorithm. The binding site was determined by 4.0 Å away from any atom of compound **1** in the EphA3 MD-IF structure. The number of energy evaluations was 1 750 000, and the number of poses was 30. Poses were clustered using all atom RMSD cutoff of 1.0 Å to remove redundancy. All other parameters were set as default.

Scoring function: Poses were further minimized by CHARMM in the rigid protein, and then sequentially filtered by three filters as defined in Figure 4. A previously reported scoring function was employed for ranking. It incorporates the hydrogen bonding penalty upon ligand binding, and uses the finite-difference Poisson approach to calculate electrostatic solvation.[21]

Biology

Phage-display-based binding assay: Experiments were performed at Ambit Biosciences Inc. (San Diego, USA) using binding assays as previously described.[17] Briefly, kinases were expressed as fusion proteins to T7 phage. T7-Kinase-tagged phage strains were mixed with known kinase inhibitors immobilized on streptavidin-coated magnetic beads and with test compounds. Test compounds that bind to the kinase ATP site displace the immobilized ligand from the kinase/phage, which is detected using quantitative polymerase chain reaction (PCR). The results are reported as the percentage of kinase/phage remaining bound to the ligand/beads, relative to a control (DMSO lacking a test compound). A small percent control (%ctrl) value indicates strong binding.

Acknowledgements

The authors thank Emilie Frugier for critical reading and English corrections to the manuscript. We are grateful to Armin Widmer (Novartis Pharma AG, Basel, Switzerland) for continuous support with the program WITNOTP, used for visual analysis. Calculations were performed on the Schroedinger cluster at the Informatikdienste of University of Zurich (Switzerland). This work was supported by the Swiss National Science Foundation (grant no. 31003A_122442 to D.H.), and by the Sino–Swiss Science and Technology Cooperation Joint Research Projects (no. IZL CZ3 123945.).

Keywords: computational chemistry · high-throughput docking · molecular dynamics · tyrosine kinases · type II kinase inhibitors

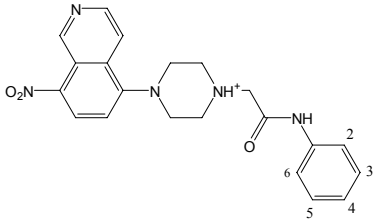
[1] O. Fedorov, S. Muller, S. Knapp, *Nat. Chem. Biol.* **2010**, *6*, 166 – 169.
[2] W. W. Chan, S. C. Wise, M. D. Kaufman, Y. M. Ahn, C. L. Ensinger, T. Haack, M. M. Hood, J. Jones, J. W. Lord, W. P. Lu, D. Miller, W. C. Patt, B. D. Smith, P. A. Petillo, T. J. Rutkoski, H. Telikepalli, L. Vogeti, T. Yao, L. Chun, R. Clark, P. Evangelista, L. C. Gavrilescu, K. Lazarides, V. M. Zaleskas, L. J. Stewart, R. A. Van Etten, D. L. Flynn, *Cancer Cell* **2011**, *19*, 556 – 568.

[3] A. Dixit, L. Yi, R. Gowthaman, A. Torkamani, N. J. Schork, G. M. Verkhivker, *PLoS One* **2009**, *4*, e7485.
[4] R. H. Adams, *Semin. Cell Dev. Biol.* **2002**, *13*, 55 – 60.
[5] R. Noberini, I. Lamberto, E. B. Pasquale, *Semin. Cell Dev. Biol.* **2012**, *23*, 51 – 57.
[6] G. Martiny-Baron, P. Holzer, E. Billy, C. Schnell, J. Brueggen, M. Ferretti, N. Schmiedeberg, J. M. Wood, P. Furet, P. Imbach, *Angiogenesis* **2010**, *13*, 259 – 267.
[7] S. A. Mitchell, M. D. Danca, P. A. Blomgren, J. W. Darrow, K. S. Currie, J. E. Kropf, S. H. Lee, S. L. Gallion, J. M. Xiong, D. A. Pippin, R. W. DeSimone, D. R. Brittelli, D. C. Eustice, A. Bourret, M. Hill-Drzewi, P. M. Maciejewski, L. L. Elkin, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6991 – 6995.
[8] Y. Liu, N. S. Gray, *Nat. Chem. Biol.* **2006**, *2*, 358 – 364.
[9] M. E. Noble, J. A. Endicott, L. N. Johnson, *Science* **2004**, *303*, 1800 – 1805.
[10] J. J. Liao, *J. Med. Chem.* **2007**, *50*, 409 – 424.
[11] F. Zuccotto, E. Ardini, E. Casale, M. Angiolini, *J. Med. Chem.* **2010**, *53*, 2681 – 2694.
[12] I. Kufareva, R. Abagyan, *J. Med. Chem.* **2008**, *51*, 7921 – 7932.
[13] M. Xu, L. Yu, B. Wan, Q. Huang, *PLoS One* **2011**, *6*, e22644.
[14] B. Nagar, W. G. Bornmann, P. Pellicena, T. Schindler, D. R. Veach, W. T. Miller, B. Clarkson, J. Kuriyan, *Cancer Res.* **2002**, *62*, 4236 – 4243.
[15] P. Bamborough, M. J. Brown, J. A. Christopher, C. W. Chung, G. W. Mellor, *J. Med. Chem.* **2011**, *54*, 5131 – 5143.
[16] Y. Choi, F. Syeda, J. R. Walker, P. J. Finerty, Jr., D. Cuerrier, A. Wojciechowski, Q. Liu, S. Dhe-Paganon, N. S. Gray, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4467 – 4470.
[17] M. W. Karaman, S. Herrgard, D. K. Treiber, P. Gallant, C. E. Atteridge, B. T. Campbell, K. W. Chan, P. Ciceri, M. I. Davis, P. T. Edeen, R. Faraoni, M. Floyd, J. P. Hunt, D. J. Lockhart, Z. V. Milanov, M. J. Morrison, G. Pallares, H. K. Patel, S. Pritchard, L. M. Wodicka, P. P. Zarrinkar, *Nat. Biotechnol.* **2008**, *26*, 127 – 132.
[18] T. L. Davis, J. R. Walker, P. Loppnau, C. Butler-Cole, A. Allali-Hassani, S. Dhe-Paganon, *Structure* **2008**, *16*, 873 – 884.
[19] C. J. A. Farenc, L. Hameetman, W. Zoutman, C. P. Tensen, G. Siegal, *FEBS Lett.* **2011**, *585*, 3593 – 3599.
[20] D. Huang, A. Caflisch, *PLoS Comput. Biol.* **2011**, *7*, e1002002.
[21] H. Zhao, D. Huang, *PLoS One* **2011**, *6*, e19923.
[22] P. Schmidtke, F. J. Luque, J. B. Murray, X. Barril, *J. Am. Chem. Soc.* **2011**, *133*, 18903 – 18910.
[23] P. Kolb, D. Huang, F. Dey, A. Caflisch, *J. Med. Chem.* **2008**, *51*, 1179 – 1188.
[24] D. Ménard, I. Niculescu-Duvaz, H. P. Dijkstra, D. Niculescu-Duvaz, B. M. Suijkerbuijk, A. Zambon, A. Nourry, E. Roman, L. Davies, H. A. Manne, F. Friedlos, R. Kirk, S. Whittaker, A. Gill, R. D. Taylor, R. Marais, C. J. Springer, *J. Med. Chem.* **2009**, *52*, 3881 – 3891.
[25] D. Niculescu-Duvaz, C. Gaulon, H. P. Dijkstra, I. Niculescu-Duvaz, A. Zambon, D. Menard, B. M. Suijkerbuijk, A. Nourry, L. Davies, H. Manne, F. Friedlos, L. Ogilvie, D. Hedley, S. Whittaker, R. Kirk, A. Gill, R. D. Taylor, F. I. Raynaud, J. Moreno-Farre, R. Marais, C. J. Springer, *J. Med. Chem.* **2009**, *52*, 2255 – 2264.
[26] B. M. Suijkerbuijk, I. Niculescu-Duvaz, C. Gaulon, H. P. Dijkstra, D. Niculescu-Duvaz, D. Menard, A. Zambon, A. Nourry, L. Davies, H. A. Manne, F. Friedlos, L. M. Ogilvie, D. Hedley, F. Lopes, N. P. Preece, J. Moreno-Farre, F. I. Raynaud, R. Kirk, S. Whittaker, R. Marais, C. J. Springer, *J. Med. Chem.* **2010**, *53*, 2741 – 2756.
[27] O. Saavedra, S. Claridge, L. Zhan, F. Raeppel, M. C. Granger, S. Raeppel, M. Mannion, F. Gaudette, N. Zhou, L. Isakovic, N. Bernstein, R. Deziel, H. Nguyen, N. Beaulieu, C. Beaulieu, I. Dupont, J. Wang, A. R. Macleod, J. M. Besterman, A. Vaisburg, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6836 – 6839.
[28] H. R. Heyman, R. R. Frey, P. F. Bousquet, G. A. Cunha, M. D. Moskey, A. A. Ahmed, N. B. Soni, P. A. Marcotte, L. J. Pease, K. B. Glaser, M. Yates, J. J. Bouska, D. H. Albert, C. L. Black-Schaefer, P. J. Dandliker, K. D. Stewart, P. Rafferty, S. K. Davidsen, M. R. Michaelides, M. L. Curtin, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1246 – 1249.
[29] W. S. Huang, X. Zhu, Y. Wang, M. Azam, D. Wen, R. Sundaramoorthi, R. M. Thomas, S. Liu, G. Banda, S. P. Lentini, S. Das, Q. Xu, J. Keats, F. Wang, S. Wardwell, Y. Ning, J. T. Snodgrass, M. I. Broudy, K. Russian, G. Q. Daley, J. Iuliucci, D. C. Dalgarno, T. Clackson, T. K. Sawyer, W. C. Shakespeare, *J. Med. Chem.* **2009**, *52*, 4743 – 4756.
[30] Y. Dai, K. Hartandi, N. B. Soni, L. J. Pease, D. R. Reuter, A. M. Olson, D. J. Osterling, S. Z. Doktor, D. H. Albert, J. J. Bouska, K. B. Glaser, P. A. Marcotte, K. D. Stewart, S. K. Davidsen, M. R. Michaelides, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 386 – 390.
[31] Z. W. Cai, D. Wei, G. M. Schroeder, L. A. Cornelius, K. Kim, X. T. Chen, R. J. Schmidt, D. K. Williams, J. S. Tokarski, Y. An, J. S. Sack, V. Manne, A. Kamath, Y. Zhang, P. Marathe, J. T. Hunt, L. J. Lombardo, J. Fargnoli, R. M. Borzilleri, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3224 – 3229.
[32] P. P. Kung, L. Funk, J. Meng, G. Alton, E. Padrique, B. Mroczkowski, *Eur. J. Med. Chem.* **2008**, *43*, 1321 – 1329.
[33] U. R. Khire, D. Bankston, J. Barbosa, D. R. Brittelli, Y. Caringal, R. Carlson, J. Dumas, T. Gane, S. L. Heald, B. Hibner, J. S. Johnson, M. E. Katz, N. Kennure, J. Kingery-Wood, W. Lee, X. G. Liu, T. B. Lowinger, I. McAlexander, M. K. Monahan, R. Natero, J. Renick, B. Riedl, H. Rong, R. N. Sibley, R. A. Smith, D. Wolanin, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 783 – 786.
[34] V. J. Cee, B. K. Albrecht, S. Geuns-Meyer, P. Hughes, S. Bellon, J. Bready, S. Caenepeel, S. C. Chaffee, A. Coxon, M. Emery, J. Fretland, P. Gallant, Y. Gu, B. L. Hodous, D. Hoffman, R. E. Johnson, R. Kendall, J. L. Kim, A. M. Long, D. McGowan, M. Morrison, P. R. Olivieri, V. F. Patel, A. Polverino, D. Powers, P. Rose, L. Wang, H. Zhao, *J. Med. Chem.* **2007**, *50*, 627 – 640.
[35] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, *Nat. Biotechnol.* **2011**, *29*, 1046 – 1051.
[36] L. M. Wodicka, P. Ciceri, M. I. Davis, J. P. Hunt, M. Floyd, S. Salerno, X. H. Hua, J. M. Ford, R. C. Armstrong, P. P. Zarrinkar, D. K. Treiber, *Chem. Biol.* **2010**, *17*, 1241 – 1249.
[37] MODELLER (copyright © 1989 – 2012 Andrej Sali): http://salilab.org/modeller/. A. Šali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779 – 815.
[38] NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign (USA): http://www.ks.uiuc.edu/Research/namd/. a) J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comp. Chem.* **2005**, *26*, 1781 – 1802; b) L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283 – 312.
[39] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586 – 3616.
[40] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926 – 935.
[41] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671 – 690.
[42] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089 – 10092.
[43] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177 – 182.
[44] F. A. Momany, R. Rone, *J. Comput. Chem.* **1992**, *13*, 888 – 900.
[45] K. T. No, J. A. Grant, H. A. Scheraga, *J. Phys. Chem.* **1990**, *94*, 4732 – 4739.
[46] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187 – 217.
[47] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545 – 1614.
[48] AutoDock 4.0 (http://autodock.scripps.edu/). a) D. S. Goodsell, A. J. Olson, *Proteins* **1990**, *8*, 195 – 202; b) G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comp. Chem.* **1998**, *19*, 1639 – 1662.

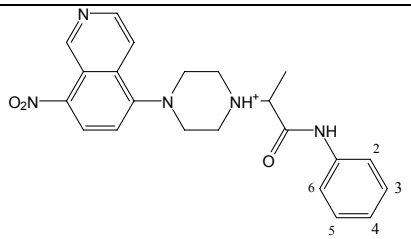
Received: June 29, 2012
Published online on ■■ ▮▮, 0000

Supporting Information

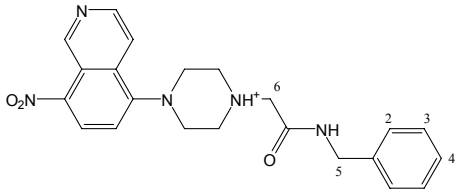
Table S1. Affinity of 5-(piperazine-1-yl)isoquinoline derivatives measured by phage display-based binding assay.[a]



	Substituents					Kinase			
	2	3	4	5	6	EphA3 wild type	Abl1 wild type	Abl1-unphosph. wild type	Abl1-unphosph. T315I
5	H	Cl	$OCHF_2$	H	H	53	>65	12	45
6	H	Br	H	H	H	63	>65	48	>65
7	H	Cl	CH_3	H	H	61	>65	48	>65
13	Cl	H	H	Cl	H	>65	>65	>65	>65
14	H	H	$N(CH_3)_2$	H	H	>65	>65	65	>65
15	F	H	H	H	H	58	>65	>65	>65
16	OCH_3	H	OCH_3	Cl	H	>65	>65	>65	>65
17	H	F	H	F	H	>65	>65	>65	>65
18	H	Cl	H	H	H	60	>65	54	>65
19	H	OCH_3	OCH_3	H	H	>65	>65	>65	>65



	2	3	4	5	6				
8[b]	H	CF_3	H	H	H	>65	>65	2.8 [3.9 µM][c]	5.8
9[b]	H	Cl	OCH_3	H	H	>65	>65	28	>65
10[b]	H	Cl	F	H	H	>65	>65	38	>65
11[b]	H	Cl	H	H	H	60	>65	21	55
20[b]	CH_3	H	H	Cl	H	>65	>65	>65	>65
21[b]	F	H	H	H	F	>65	>65	>65	>65
22[b]	H	CH_3	H	H	OCH_3	>65	>65	>65	>65



	2	3	4	5	6				
23[b]	Cl	H	H	H	CH_3	>65	>65	>65	>65
24[b]	CH_3	H	H	CH_3	H	61	>65	>65	>65
25	CH_3	H	H	H	H	>65	>65	>65	>65
26	OCH_3	H	H	H	H	>65	>65	>65	>65
27	OCF_3	H	H	H	H	>65	>65	>65	>65
28[b]	OCH_3	H	H	H	CH_3	>65	>65	>65	>65

[a]Percentage of kinase not displaced by the test compounds at 30 μM concentration with small values indicating high affinity. [b]Compounds are racemic mixtures as purchased, and checked by in-house ESI-MS. [c]The K_d value in parentheses is mean of two dose response measurements of 11 points each with K_d of 3.8 μM and 4.0 μM, respectively.

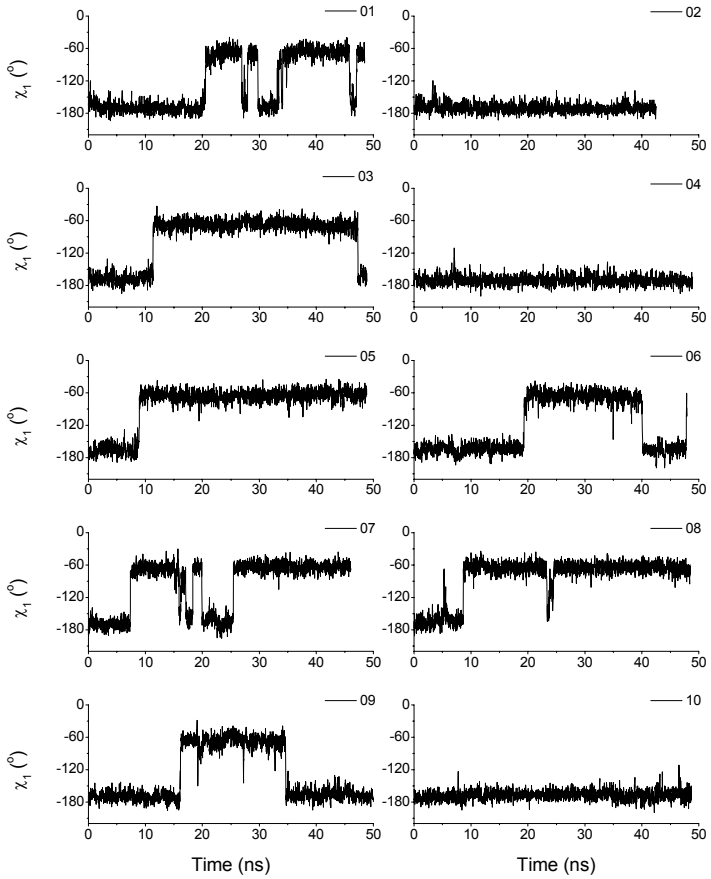


Figure S1. Time evolution of the dihedral angle χ_1 of Tyr742 in 10 explicit solvent MD simulations of EphA3 (3DZQ), in the presence of trifluoromethyl-benzene in the allosteric site. The χ_1 value around -175° corresponds to the orientation in the crystal structure (3DZQ). The percentage of snapshots with χ_1 close to -175 and -60 is 57% and 43%, respectively. Note that the variability in the number of transitions in different runs is a consequence of the stochastic character of the side chain rotation which is governed by an activation energy barrier.

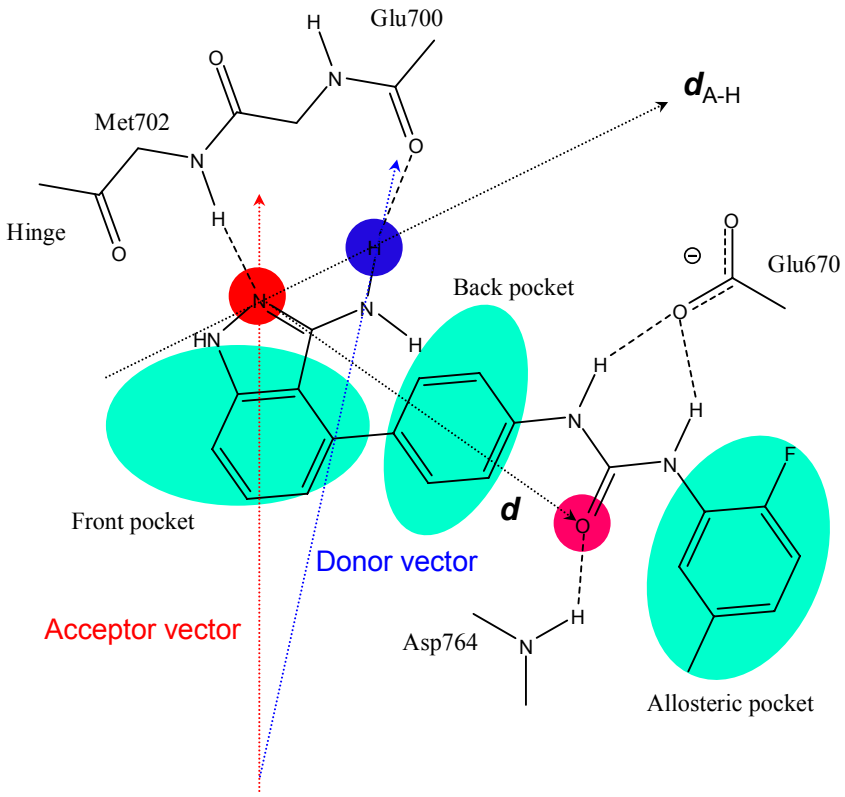


Figure S2. Illustration of the geometric constraints in pharmacophore tailoring ZINC compounds library. Donor vector is defined as from the donor heavy atom to the attached hydrogen atom. Acceptor vector is defined as that bisects the angle centered at the acceptor and points to the lone pair or from carbon to oxygen in the case of carbonyl oxygen. Pharmacophore geometric constraints based on known type II inhibitors and the geometry of the binding site are:

1. Capability to form bi-dentate hydrogen bond with the hinge. The angle between the donor and acceptor vector is less than 90 degree, and the distance between the donor hydrogen and the acceptor is less than 2.9 Å.
2. In case of forming bi-dentate hydrogen bond, acceptor-donor hydrogen vector $\boldsymbol{d}_{A-H}$ is defined as from the acceptor to the donor hydrogen.
3. The centroid of each hydrophobic ring is constrained by $|\boldsymbol{d}|$ and $\boldsymbol{d}\cdot\boldsymbol{d}_{A-H}/|\boldsymbol{d}_{A-H}|$, wherein $\boldsymbol{d}$ is a vector from the acceptor (red circle) to the centroid. Specifically, for ring in the front site $|\boldsymbol{d}|<4.5$ Å and $|\boldsymbol{d}\cdot\boldsymbol{d}_{A-H}|\cdot|\boldsymbol{d}_{A-H}|^{-1}<2$ Å; for the ring in the back site 4 Å$<|\boldsymbol{d}|<10$ Å and $\boldsymbol{d}\cdot\boldsymbol{d}_{A-H}>0$; for the ring in the allosteric site $|\boldsymbol{d}|>11$ Å and $\boldsymbol{d}\cdot\boldsymbol{d}_{A-H}>0$;
4. For the additional acceptor (pink circle) 6 Å$<|\boldsymbol{d}|<11$ Å and $\boldsymbol{d}\cdot\boldsymbol{d}_{A-H}>0$.

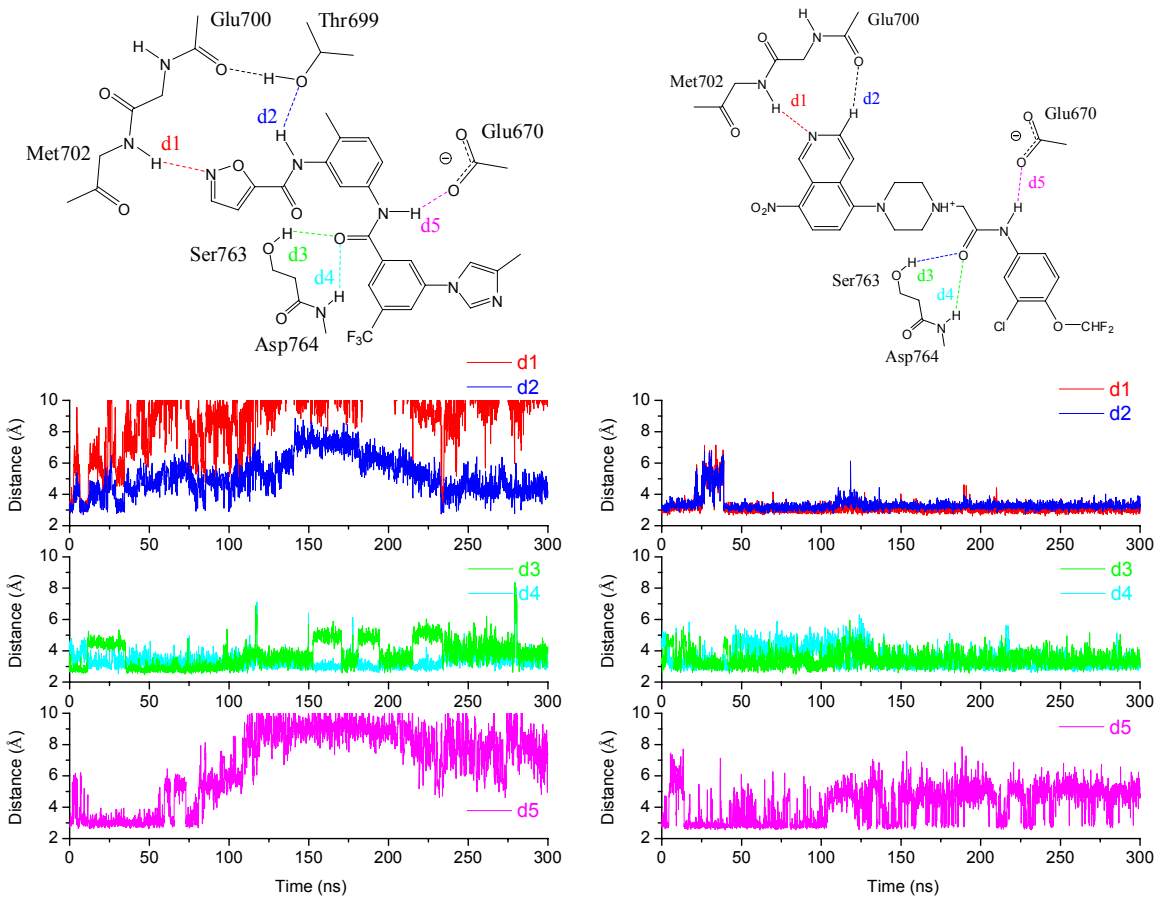


Figure S3. Explicit solvent MD validation of binding mode of compound 5. (Left) MD run started from the X-ray structure of the complex between EphA3 and inhibitor **1** (PDB code 3DZQ). (Right) MD run started from the docked pose of compound **5** into the MD-IF structure of EphA3. (Top) Two-dimensional illustrations; (Bottom) time evolutions of hydrogen bonding distances measured between donor and acceptor atoms. Note that values around 5 Å reflect single water bridged hydrogen bonds with Glu670. Note also that these runs used different random numbers for the generation of initial set of velocities than those in Figure 6 of the main text.

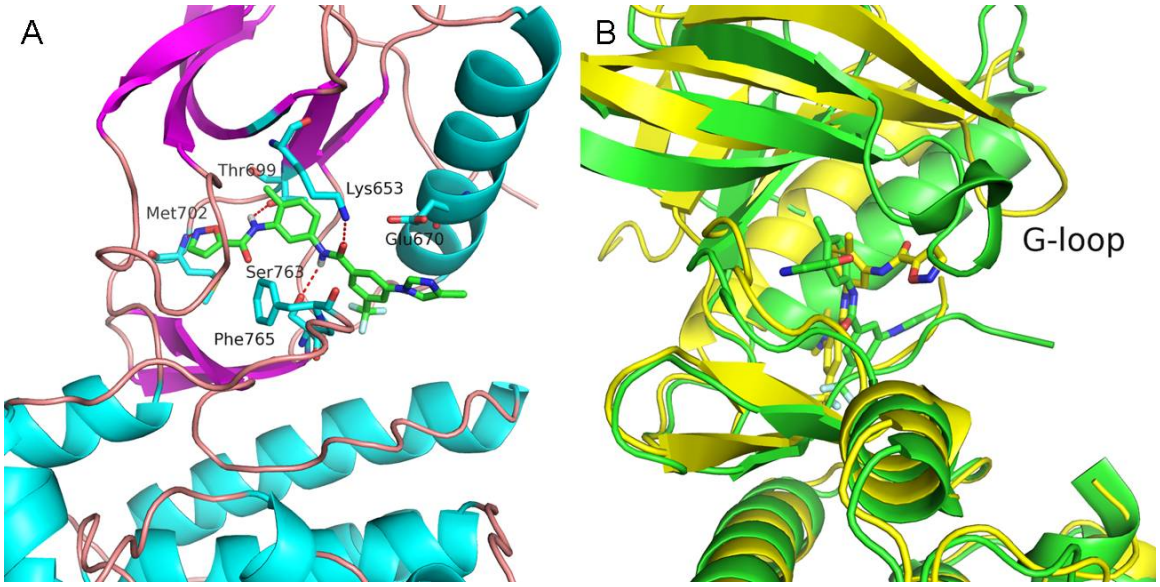


Figure S4. Snapshots from two MD simulations with inhibitor 1. Snapshot from the simulation in Figure 6 at about 230 ns (A) and from Figure S3 at 150 ns (yellow) with the one in X-ray structure (green). In A, the linker amide group was flipped by 180°. In B, the G-loop showed significant conformational change compared to the X-ray structure.

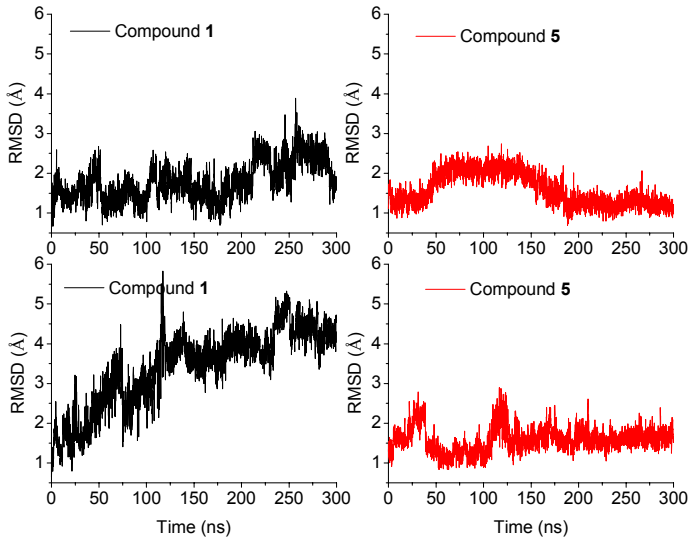


Figure S5. Time evolutions of the heavy atom root-mean-square deviation (RMSD) of compound 1 (left) and 5 (right) upon overlapping the binding site C_α atoms to the EphA3 X-ray structure and MD-IF structure, respectively. In one of the two simulations with compound 1 (left bottom, also shown in Figures S3 and S4), the G-loop is displaced with respect to the hinge and the isoxazole group of compound 1 moves away from the hinge after about 50 ns.

Chapter 4

Discovery of a Novel Chemotype of Tyrosine Kinase Inhibitors by Fragment-Based Docking and Molecular Dynamics

ACS Medicinal Chemistry Letters
Letter
pubs.acs.org/acsmedchemlett

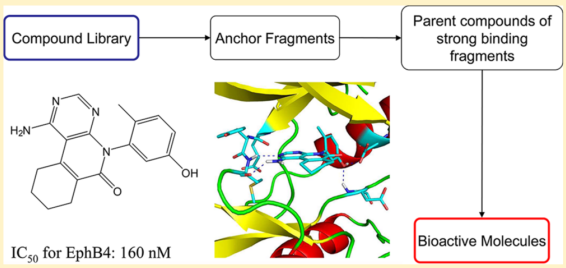
Discovery of a Novel Chemotype of Tyrosine Kinase Inhibitors by Fragment-Based Docking and Molecular Dynamics

Hongtao Zhao,[†] Jing Dong,[†] Karine Lafleur,[‡] Cristina Nevado,*[,‡] and Amedeo Caflisch*[,†]

[†]Department of Biochemistry and [‡]Department of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Supporting Information

ABSTRACT: We have discovered a novel chemical class of inhibitors of the EphB4 tyrosine kinase by fragment-based high-throughput docking followed by explicit solvent molecular dynamics simulations for assessment of the binding mode. The synthesis of a single derivative (compound 7) of the hit identified in silico has resulted in an improvement of the inhibitory potency in an enzymatic assay from 8.4 μM to 160 nM and a ligand efficiency of 0.39 kcal/mol per non-hydrogen atom. Such remarkable improvement in affinity is due to an additional hydroxyl group involved in two favorable (buried) hydrogen bonds as predicted by molecular dynamics and validated by the crystal structure of the complex with EphA3 solved at 1.7 Å resolution.



KEYWORDS: *In silico screening, EphB4 kinase, angiogenesis, cancer, explicit solvent MD*

The Eph-ephrin system, including the EphA2 and EphB4 receptors, plays a critical role in tumor and vascular functions during carcinogenesis.[1,2] Recently, it has been shown that delivery of chemotherapeutic drugs by an EphA2 targeting peptide into EphA2-expressing cancer cells led to dramatically improved efficacy in inhibiting tumor growth.[3] So far, a few Eph inhibitors have been identified, including the marketed drug Dasatinib (Figure S1 in the Supporting Information).[4−12] Although their role is still controversial for certain types of cancer, e.g., non small cell lung cancer,[13] the identification of selective inhibitors of Eph tyrosine kinases will help to elucidate their involvement in deregulated signaling.

Previously, we have developed an efficient in silico procedure called ALTA, which stays for anchor-based library tailoring approach, to interrogate a library of compounds for high-throughput docking.[14] First, small and mainly rigid virtual fragments are docked in the binding site. The fragments with most favorable calculated binding free energy (anchors) are used to identify the compounds with 2D structure containing one of these anchors, which are then submitted to flexible-ligand docking. In this letter, we report a new approach for in silico screening based on the synergistic combination of the ALTA procedure for docking followed by explicit solvent molecular dynamics simulations for further validation of the binding poses.

The flowchart of the ALTA procedure is shown in Figure 1. First, the nearly 9 million compounds in the ZINC-all now library[15] (version of August 2011) were decomposed into 563,774 fragments by in house developed software (Figure S2 in the Supporting Information). Just like its in vitro counterpart of fragment-based drug discovery,[16,17] the success of the ALTA

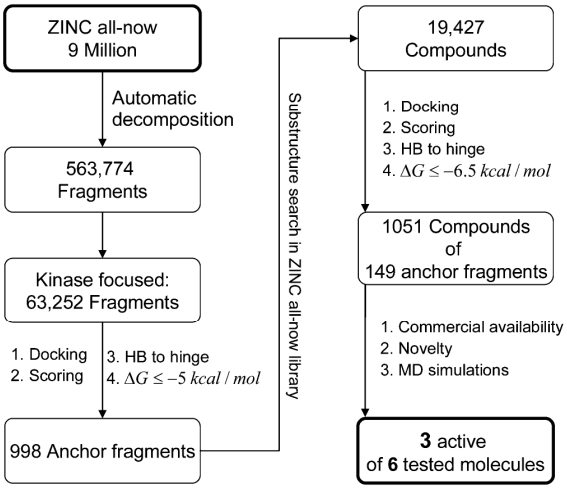
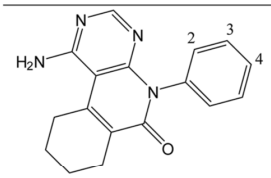
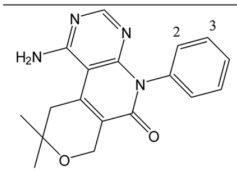
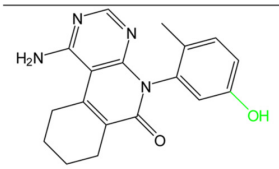


Figure 1. Flowchart of the ALTA virtual screening approach for the tyrosine kinase EphB4.

in silico screening approach depends on the choice of fragments. The use of virtual fragments by computational decomposition of a real compound library offers opportunities to explore a much greater fragmental space, with no limitations in availability. To obtain fragments with high chemical richness

Received: July 16, 2012
Accepted: August 23, 2012
Published: August 23, 2012

Table 1. Novel EphB4 Kinase Inhibitors Discovered by the ALTA Virtual Screening Approach with Hit Optimization Colored in Green

Compounds		Substituents	IC_{50} (μM)[a]	IC_{50} (μM)[b]	LE[c]
	**1**	2-methyl	48%[d]	8.4	0.30
	2	3-methoxy	9.4 (75%[d])	5.2	0.30
	3	4-methoxy	37%[d]	N.A.[e]	N.A.
	4	2-methoxy	5%[d]	N.A.	N.A.
	**5**	3-methoxy	38%[d]	14.3	0.25
	6	2-methoxy	0%[d]	N.A.	N.A.
	**7**		0.30	0.16	0.39

[a]Enzymatic assay based on FRET carried out in house using Invitrogen EphB4 assay kit at an ATP concentration of 30 μM. [b]Enzymatic assay with radioactive ATP at 1 μM concentration carried out at Reaction Biology Corp. The IC_{50} values were determined in a 10-dose response. [c]Ligand efficiency in kcal/mol per heavy atom. [d]Percentage inhibition at compound concentration of 30 μM and ATP concentration of 30 μM. [e]N.A.: not available.

that can serve as a starting point either directly for hit optimization or for identification of their "parent" compounds, we developed a new decomposition protocol whose main difference from our previous approach[18] is the preservation of longer substituents (e.g., *N*-methylurea) on ring systems (details of algorithm explained in Figure S2 in the Supporting Information). Most of the fragments obtained by the new decomposition algorithm have a molecular weight ranging between 150 and 300 g/mol, possess fewer than five rotatable bonds, and do not have any formal charge (Figure S3 in the Supporting Information). Second, this set of fragments was reduced to a kinase-focused library of 63,252 fragments by retaining only those with molecular weight smaller than 300 g/mol, a maximum of three rotatable bonds, more than one ring, and the capability to form two hydrogen bonds with the backbone polar groups of the so-called hinge region. For the latter criterion acidic CH groups (e.g., in aromatic rings) were also considered as donors. The requirement of having more than one ring helps to direct the search toward chemical space less affected by the crowded intellectual property coverage, given the diversity in fused rings. Moreover, one-ring anchor fragments are too small for providing enough binding energy. Use of a target-focused fragment library is computationally more efficient than docking the entire library of fragments.

The kinase-focused fragment library was then docked into the ATP binding site of EphB4 (PDB code 2VWX) by AutoDock4,[19] followed by ranking according to a previously reported scoring function.[8] The 998 fragments with an estimated binding energy lower than −5 kcal/mol were used to identify their "parent" compounds in the ZINC-all now library which yielded a total of 19,427 compounds. Flexible-ligand docking of these compounds followed by scoring[8] and elimination of those mentioned in patents related to kinases resulted in four scaffolds, which were further investigated by explicit solvent molecular dynamics simulations using the all-atom CHARMM PARAM22 force field[20] and the TIP3P model of water.[21] Two molecules sharing the pyrimidoisoquinolinone scaffold showed stable intermolecular hydrogen bonds with the hinge region (bottom panels of Figure S4 in the Supporting Information). Thus, commercially available compounds **1**−**6**, bearing a pyrimidoisoquinolinone scaffold, were purchased and tested (Table 1). Of these six compounds, **1**, **2**, and **5** showed inhibition of the EphB4 tyrosine kinase in the low micromolar range in two different enzymatic assays based on FRET (carried out in house) and radioactive ATP (performed at Reaction Biology Corp.), respectively.

The previously reported optimization of another chemical class of EphB4 inhibitors discovered by the ALTA approach[7,14] suggested to us that the addition of a hydroxyl group in position 5 of the phenyl ring in compound **1** would increase the binding affinity. This hydroxyl group is expected to be involved in two additional hydrogen bonds with the side chain of Glu664 and the backbone NH of the Asp in the DFG motif, as observed in so-called type I1/2 inhibitors[22] of tyrosine kinase with small gatekeeper residue (e.g., Thr) and DFG-in conformation. As predicted, compound **7** shows four stable hydrogen bonds with the ATP binding site (two with the hinge region and two involving the additional hydroxyl group) in two 300-ns molecular dynamics simulations with explicit solvent (Figure 2). Note that these molecular dynamics runs were started from the binding mode obtained by docking and were carried out before chemical synthesis of compound **7**.

The synthesis of compound **7** started with the saponification of ethyl 2-oxocyclohexanecarboxylate **8** in the presence of sodium hydroxide. Reaction of the β-keto acid **9** with acetone in the presence of acetic anhydride and concentrated sulfuric

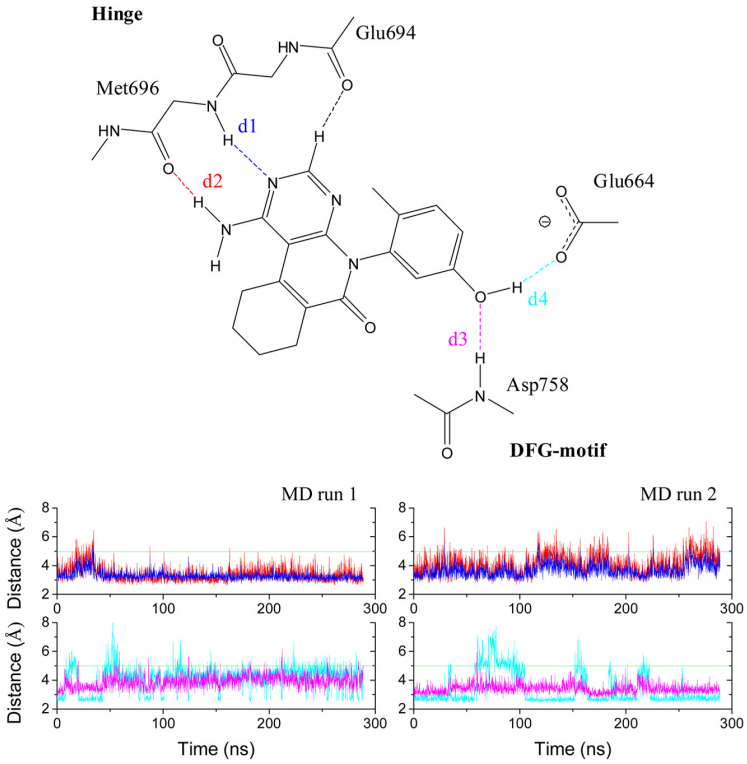
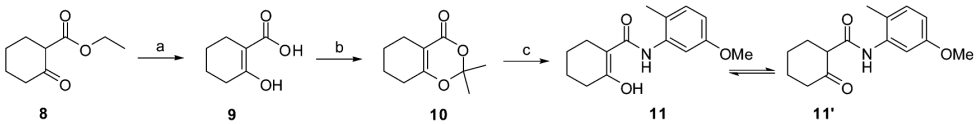


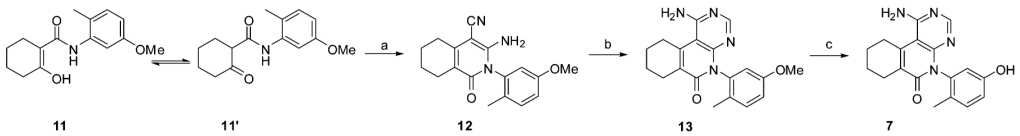
Figure 2. 2D illustration of binding mode of inhibitor **7** in the ATP binding site of EphB4 (top) and time series of intermolecular hydrogen bonds, i.e., distances between donor and acceptor atoms, in two molecular dynamics simulations started with different random values of the initial atomic velocities (bottom). The choice of colors is consistent in top and bottom panels while the green horizontal lines in the latter are drawn to illustrate that the hydrogen bonds are always formed except for transient ruptures.

Scheme 1. Synthesis of Tautomers 11 and 11′[a]



[a]Reagents and conditions: (a) NaOH, water, 25 °C, 12 h, 61%; (b) acetone, H_2SO_4, acetic anhydride, 0 °C, 4 h, 93%; (c) 5-methoxy-2-methylaniline, *o*-xylene, MW, 150 °C, 5 min, 46%.

Scheme 2. Synthesis of Target Compound 7[a]



[a]Reagents and conditions: (a) malononitrile, piperidine, EtOH, 100 °C, 2 h, 82%; (b) formamide, 210 °C, 7 h, 63%; (c) BBr_3, CH_2Cl_2, 25 °C, 16 h, 28%.

acid provided the activated acetal **10**,[23] which reacted with 5-methoxy-2-methylaniline under microwave irradiation to afford a tautomeric mixture of amides **11** and **11′** (Scheme 1).[24] Reaction of the mixture with malononitrile provided intermediate **12**, which upon condensation with formamide afforded compound **13**.[25] Final demethylation for the methoxy group in the aromatic ring was obtained in the presence of BBr_3 (Scheme 2).

It was gratifying to observe that in the enzymatic assay the potency of compound **7** is about 50 times higher than that of the parent compound **1** (Table 1). This improvement corresponds to a difference in binding free energy of about 2.5 kcal/mol, which is consistent with the two additional hydrogen bonds of the hydroxyl group of **7** as predicted by the docking and molecular dynamics simulations. The final validation was obtained by the determination of the crystal

dx.doi.org/10.1021/ml3001984 | *ACS Med. Chem. Lett.* 2012, 3, 834−838

structure of the complex between the catalytic domain of EphA3 and inhibitor 7 (resolution of 1.7 Å, see Supporting Information for details of X-ray crystallography). The binding mode in the crystal structure confirmed the predicted binding pose of inhibitor 7 into EphB4 (DFG-in conformation) and the four intermolecular hydrogen bonds observed in the molecular dynamics runs (Figure 3). The EphA3 tyrosine kinase was used for determination of the crystal structure because of its straightforward expression in *Escherichia coli*[10] while expression of EphB4 has been reported only in insect cells.[5] Note that 32 of the 36 residues in the ATP binding site of EphA3 are identical to those of EphB4.[26] Moreover, the side chains involved in binding compound 7 are identical in EphA3 and EphB4.

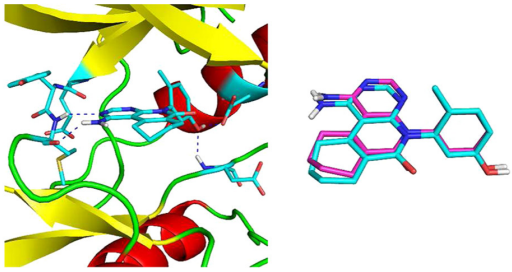


Figure 3. (Left) Crystal structure of inhibitor 7 bound to the kinase domain of EphA3. Dashed lines represent intermolecular hydrogen bonds. (Right) Superposition of pose of 7 in EphB4 obtained by docking and molecular dynamics (carbon atoms in magenta) onto the pose of 7 in EphA3 obtained by X-ray crystallography (carbon atoms in cyan). The overlap was obtained by structural alignment of the crystal structure of the EphA3/inhibitor 7 complex to the in silico predicted complex of 7 and EphB4 using only the C_α atoms of the two kinase domains.

The affinity of compound 7 was further tested on a panel of five tyrosine kinases, known as validated drug targets in cancer therapy. The IC_{50} value in enzymatic assay in the presence of the recombinant catalytic domain and 1 μM radiolabeled ATP is 0.338 μM for Src, 0.864 μM for Abl1, 1.38 μM for Lck, 1.62 μM for EGFR, while no inhibition was observed for IGF1R. Thus compound 7 has higher affinity for EphB4 than for these five tyrosine kinases.

Compound 7 was also tested in a cell-based assay. The cellular EphB4 phosphorylation assay (performed at ProQinase GmbH) makes use of the murine embryonal fibroblast cell line (MEF), which expresses a high level of exogenously introduced full-length human EphB4. Stimulation of these cells with 2 μg/mL soluble human Ephrin-B2/Fc chimera for 2 h results in receptor tyrosine autophosphorylation. Compound 7 shows a cellular IC_{50} of 0.93 μM, indicating its potential for future optimization.

In conclusion, we have discovered a new chemical class of kinase inhibitors by fragment-based docking into the ATP-binding site of the EphB4 kinase and molecular dynamics simulations for further assessment of the binding mode. The chemical synthesis of a single derivative (compound 7) by addition of just one hydroxyl group, designed on the basis of the simulation results, has resulted in a novel chemotype of nanomolar inhibitors of the EphB4 tyrosine kinase. Finally, the determination of the crystal structure of the complex of EphA3 with compound 7 has definitively validated the binding mode predicted in silico.

■ ASSOCIATED CONTENT

Supporting Information

Chemical synthesis, crystallization, similarity comparison with known Eph inhibitors, description of the algorithm used to decompose a compound library, distribution of physicochemical properties of the library of anchor fragments, and MD screening of candidate compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The atomic coordinates and structure factors of EphA3 in complex with compound 7 have been deposited with the Protein Data Bank as entry 4G2F.

■ AUTHOR INFORMATION

Corresponding Author

*C.N.: phone, (41) 446353945; fax, (41) 446353948; e-mail, nevado@oci.uzh.ch. A.C.: phone, (41) 446355521; fax, (41) 446356862; e-mail, caflisch@bioc.uzh.ch.

Funding

This work was supported by grants of the Swiss National Science Foundation and the Sino-Swiss Science and Technology Cooperation to A.C.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Danzhi Huang for interesting discussions and Mr. Tim Knehans for the help on calculation of similarity index. We thank Xiao-Dan Li and Chitra Rajendran (Paul Scherrer Institute) for interesting discussions and help with X-ray data collection in a preliminary phase of this work. The simulations were carried out on the Schrödinger cluster at the University of Zurich.

■ ABBREVIATIONS

EphB4, erythropoietin producing human hepatocellular carcinoma receptor B4; ALTA, anchor-based library tailoring approach; MD, molecular dynamics

■ REFERENCES

(1) Noberini, R.; Lamberto, I.; Pasquale, E. B. Targeting Eph receptors with peptides and small molecules: progress and challenges. *Semin. Cell Dev. Biol.* **2012**, *23*, 51−7.

(2) Tognolini, M.; Incerti, M.; Hassan-Mohamed, I.; Giorgio, C.; Russo, S.; Bruni, R.; Lelli, B.; Bracci, L.; Noberini, R.; Pasquale, E. B.; Barocelli, E.; Vicini, P.; Mor, M.; Lodola, A. Structure-activity relationships and mechanism of action of Eph-ephrin antagonists: interaction of cholanic acid with the EphA2 receptor. *ChemMedChem* **2012**, *7*, 1071−83.

(3) Wang, S.; Placzek, W. J.; Stebbins, J. L.; Mitra, S.; Noberini, R.; Koolpe, M.; Zhang, Z.; Dahl, R.; Pasquale, E. B.; Pellecchia, M. Novel targeted system to deliver chemotherapeutic drugs to EphA2-expressing cancer cells. *J. Med. Chem.* **2012**, *55*, 2427−36.

(4) Miyazaki, Y.; Nakano, M.; Sato, H.; Truesdale, A. T.; Stuart, J. D.; Nartey, E. N.; Hightower, K. E.; Kane-Carson, L. Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro evaluation targeting angiogenetic kinases. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 250−4.

(5) Bardelle, C.; Cross, D.; Davenport, S.; Kettle, J. G.; Ko, E. J.; Leach, A. G.; Mortlock, A.; Read, J.; Roberts, N. J.; Robins, P.;

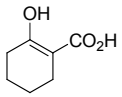
Williams, E. J. Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2776−80.
(6) Sun, L.; Liang, C.; Shirazian, S.; Zhou, Y.; Miller, T.; Cui, J.; Fukuda, J. Y.; Chu, J. Y.; Nematalla, A.; Wang, X.; Chen, H.; Sistla, A.; Luu, T. C.; Tang, F.; Wei, J.; Tang, C. Discovery of 5-[5-fluoro-2-oxo-1,2- dihydroindol-(3Z)-ylidenemethyl]-2,4- dimethyl-1H-pyrrole-3-carboxylic acid (2-diethylaminoethyl)amide, a novel tyrosine kinase inhibitor targeting vascular endothelial and platelet-derived growth factor receptor tyrosine kinase. *J. Med. Chem.* **2003**, *46*, 1116−9.
(7) Lafleur, K.; Huang, D.; Zhou, T.; Caflisch, A.; Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **2009**, *52*, 6433−46.
(8) Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011**, *6*, e19923.
(9) van Linden, O. P.; Farenc, C.; Zoutman, W. H.; Hameetman, L.; Wijtmans, M.; Leurs, R.; Tensen, C. P.; Siegal, G.; de Esch, I. J. Fragment based lead discovery of small molecule inhibitors for the EPHA4 receptor tyrosine kinase. *Eur. J. Med. Chem.* **2012**, *47*, 493−500.
(10) Choi, Y.; Syeda, F.; Walker, J. R.; Finerty, P. J., Jr.; Cuerrier, D.; Wojciechowski, A.; Liu, Q.; Dhe-Paganon, S.; Gray, N. S. Discovery and structural analysis of Eph receptor tyrosine kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4467−70.
(11) Martiny-Baron, G.; Holzer, P.; Billy, E.; Schnell, C.; Brueggen, J.; Ferretti, M.; Schmiedeberg, N.; Wood, J. M.; Furet, P.; Imbach, P. The small molecule specific EphB4 kinase inhibitor NVP-BHG712 inhibits VEGF driven angiogenesis. *Angiogenesis* **2010**, *13*, 259−67.
(12) Zhao, H.; Huang, D.; Caflisch, A. Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. *ChemMedChem* **2012**, DOI: 10.1002/cmdc.201200331.
(13) Li, G.; Ji, X. D.; Gao, H.; Zhao, J. S.; Xu, J. F.; Sun, Z. J.; Deng, Y. Z.; Shi, S.; Feng, Y. X.; Zhu, Y. Q.; Wang, T.; Li, J. J.; Xie, D. EphB3 suppresses non-small-cell lung cancer metastasis via a PP2A/RACK1/Akt signalling complex. *Nat. Commun.* **2012**, *3*, 667.
(14) Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73*, 11−8.
(15) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−82.
(16) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent developments in fragment-based drug discovery. *J. Med. Chem.* **2008**, *51*, 3661−80.
(17) Roughley, S. D.; Hubbard, R. E. How well can fragments explore accessed chemical space? A case study from heat shock protein 90. *J. Med. Chem.* **2011**, *54*, 3989−4005.
(18) Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **2006**, *49*, 7384−92.
(19) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195−202.
(20) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−614.
(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(22) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681−94.
(23) Sato, M.; Ogasawara, H.; Oi, K.; Kato, T. Synthesis of 1,3-Dioxin-4-One Derivatives. *Chem. Pharm. Bull.* **1983**, *31*, 1896−1901.
(24) Miriyala, B.; Williamson, J. S. An efficient and rapid synthesis of beta-carboxamide derivatives using 2,2-dimethyl-2H,4H-1,3-dioxin-4-ones by microwave irradiation. *Tetrahedron Lett.* **2003**, *44*, 7957−7959.
(25) Bogdanowiczszwed, K.; Policht, A. The Reaction of Malononitrile with Enamines of Some Cyclic Beta-Keto Acid Anilides and Amidines. *J. Prakt. Chem.* **1984**, *326*, 721−728.
(26) Huang, D.; Zhou, T.; Lafleur, K.; Nevado, C.; Caflisch, A. Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. *Bioinformatics* **2010**, *26*, 198−204.

Supporting Information

Synthesis of Compound 7

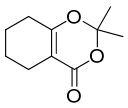
All reagents were used as received unless otherwise noted. Solvents were purchased in the best quality available, degassed by purging thoroughly with nitrogen and dried over activated molecular sieves of appropriate size. Alternatively, they were purged with argon and passed through alumina columns in a solvent purification system (Innovative Technology). Reactions were monitored by thin layer chromatography (TLC) using Merck TLC silica gel 60 F254. Flash column chromatography was performed over silica gel (230-400 mesh). NMR spectra were recorded on AV2 400 or AV2 500 MHz Bruker spectrometers. Chemical shifts are given in ppm. The spectra are calibrated to the residual ^{1}H and ^{13}C signals of the solvents. Multiplicities are abbreviated as follows: singlet (s), doublet (d), triplet (t), quartet (q), doublet-doublet (dd), quintet (quint), septet (sept), multiplet (m), and broad (br). Melting points were determined on a Büchi Melting Point B-540 instrument. The purity of all tested compounds was determined by HPLC on a Waters Acquity UPLC (Waters, Milford, USA) Top spectrometer using an Acquity BEH C18 HPLC column (1.7 µm, 1x50 mm, Waters) with a mixture of H_2O + 0.1% HCOOH (A) and CH_3CN + 0.1% HCOOH (B) solvent (0.1 ml flow rate, linear gradient from 5 to 98% B within 4 min followed by flushing with 98% B for 1 min). UV detection was set to 200-260 nm.

2-Hydroxycyclohex-1-enecarboxylic acid (9)



Ethyl 2-oxocyclohexanecarboxylate (1 mL, 6.3 mmol) was added to a solution of NaOH (275 mg, 6.9 mmol) in water (12 mL) at 0 °C. The resulting reaction mixture was stirred under ice-cooling for 3 h, and at room temperature for 12 h. The aqueous solution was washed with ether and acidified with concentrated HCl under ice cooling. The mixture was stirred for 45 min, and the formed precipitate was filtered off and dried to afford the expected product as a white solid (549 mg, 61%). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 11.98 (s, 1H), 2.32-2.25 (m, 4H), 1.75-1.59 (m, 4H), OH not observed; ^{13}C NMR (100 MHz, $CDCl_3$): δ = 176.3, 174.7, 97.1, 29.3, 22.4, 22.3, 21.8; IR (film): $\tilde{\upsilon}$ = 3005, 2940, 2860, 1639, 1584, 1434, 1210, 795 cm^{-1}; MS (ESI): *m/z*: calcd for $C_7H_9O_3^-$: 141.2, found: 140.9.

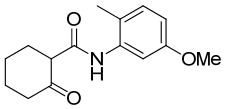
2,2-Dimethyl-5,6,7,8-tetrahydro-4H-benzo[d][1,3]dioxin-4-one (10)



To a solution of 2-hydroxycyclohex-1-enecarboxylic acid (406 mg, 2.8 mmol), acetone (417 µL, 5.6 mmol) and acetic anhydride (582 µL, 6.1 mmol) at -5 °C was added concentrated sulfuric acid (36 µL, 0.7 mmol). The reaction was stirred at 0 °C for 4 h, poured

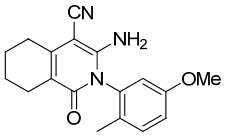
into a 10% solution of Na_2CO_3, and extracted with ether. The organic layer was dried over $MgSO_4$, filtered, and concentrated under reduced pressure to afford the desired product as a colorless solid (484 mg, 93%). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 2.29 (tt, J = 6.2 Hz, J = 1.8 Hz, 2H), 2.18 (tt, J = 6.3 Hz, J = 1.8 Hz, 2H), 1.76-1.71 (m, 2H), 1.69-1.66 (m, 8H); ^{13}C NMR (100 MHz, $CDCl_3$): δ = 164.8, 162.0, 105.1, 102.2, 27.4, 25.1, 21.9, 21.6, 21.1; IR (film): $\tilde{\upsilon}$ = 2996, 2939, 2859, 1719, 1649, 1401, 1296, 1202, 1024, 760 cm^{-1}; MS (ESI): *m/z*: calcd for $C_{10}H_{14}O_3Na^+$: 205.1, found: 205.0.

_N_-(5-Methoxy-2-methylphenyl)-2-oxocyclohexanecarboxamide (11)



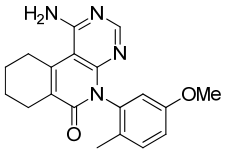
A mixture of 2,2-dimethyl-5,6,7,8-tetrahydro-4H-benzo[d][1,3]dioxin-4-one (457 mg, 2.5 mmol) and 5-methoxy-2-methylaniline (343 mg, 2.5 mmol) in o-xylene (8 mL) was subjected to microwave irradiation for 5 min at 150 °C. The reaction mixture was cooled down, and the resulting precipitate removed by filtration over celite. The mixture was extracted with EtOAc and washed with HCl 10%. The organic layer was dried over $MgSO_4$, filtered, and concentrated under reduced pressure. Purification by column chromatography (gradient toluene/EtOAc 10:0 to 10:1) afforded the desired product as a white solid (308 mg, 46%). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 9.41 (s, 1H), 7.73 (d, J = 2.6 Hz, 1H), 7.06 (d, J = 8.4 Hz, 1H), 6.61 (dd, J = 8.4 Hz, J = 2.6 Hz, 1H), 3.79 (s, 3H), 3.35 (ddd, J = 11.1 Hz, J = 5.5 Hz, J = 0.9 Hz, 1H), 2.66-2.63 (m, 1H), 2.54-2.43 (m, 2H), 2.27 (s, 3H), 2.14-2.10 (m, 1H), 2.05-1.93 (m, 2H), 1.87-1.77 (m, 2H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 212.0, 167.0, 158.3, 136.8, 130.8, 119.9, 110.7, 107.3, 55.6, 55.4, 42.5, 32.4, 27.6, 24.7, 17.0; IR (film): $\tilde{\upsilon}$ = 3227, 3122, 2939, 1709, 1496, 1281, 1122, 1037, 803 cm^{-1}; HRMS (ESI): *m/z*: calcd for $C_{15}H_{19}NO_3Na^+$: 284.1257, found: 284.1257.

3-Amino-2-(5-methoxy-2-methylphenyl)-1-oxo-1,2,5,6,7,8-hexahydroisoquinoline-4-carbonitrile (12)



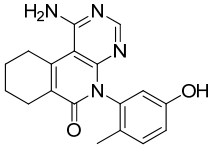
A mixture of N-(5-methoxy-2-methylphenyl)-2-oxocyclohexanecarboxamide (280 mg, 1.1 mmol), malononitrile (141 mg, 2.1 mmol), piperidine (63 µL, 0.6 mmol) in EtOH (3 mL) was heated to 100 °C for 2 h. The reaction mixture was cooled down, and the resulting precipitate was filtered off, affording the expected product as a white solid (273 mg, 82%). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.32 (d, J = 8.5 Hz, 1H), 6.98 (dd, J = 8.5 Hz, J = 2.3 Hz, 1H), 6.68 (d, J = 2.3 Hz, 1H), 4.67 (s, 2H), 3.80 (s, 3H), 2.67-2.65 (m, 2H), 2.45-2.43 (m, 2H), 2.04 (s, 3H), 1.81-1.73 (m, 4H); ^{13}C NMR (125 MHz, $CDCl_3$): δ = 160.7, 159.3, 152.3, 147.1, 133.8, 132.7, 127.8, 117.2, 116.5, 116.2, 113.1, 72.1, 55.5, 28.2, 22.9, 22.0, 21.8, 16.3; IR (film): $\tilde{\upsilon}$ = 3419, 3308, 3208, 2869, 2202, 1662, 1606, 1531, 1448, 1242, 842 cm^{-1}; HRMS (ESI): *m/z*: calcd for $C_{18}H_{19}N_3O_2Na^+$: 332.1370, found: 332.1370.

1-Amino-5-(5-methoxy-2-methylphenyl)-7,8,9,10-tetrahydropyrimido[4,5-c]isoquinolin-6(5H)-one (13)



3-Amino-2-(5-methoxy-2-methylphenyl)-1-oxo-1,2,5,6,7,8-hexahydroisoquinoline-4-carbonitrile (260 mg, 0.8 mmol) was dissolved in formamide (2.6 mL), and the resulting solution was heated to 210 °C for 7 h. The reaction was poured into ice water, and the resulting solid was filtered off. Purification by column chromatography (gradient EtOAc/CH_2Cl_2 1:1 to 2:1) afforded the expected product as a white solid (180 mg, 63%). ^{1}H NMR (500 MHz, DMSO-d_6): δ = 8.02 (s, 1H), 7.23 (d, J = 8.4 Hz, 1H), 7.11 (s, 2H), 6.89 (dd, J = 8.4 Hz, J = 2.6 Hz, 1H), 6.65 (d, J = 2.6 Hz, 1H), 3.71 (s, 3H), 3.04-3.03 (m, 2H), 2.47-2.46 (m, 2H), 1.76 (s, 3H), 1.73-1.70 (m, 4H); ^{13}C NMR (125 MHz, DMSO-d_6): δ = 161.0, 160.5, 158.0, 155.7, 153.7, 143.1, 137.9, 130.7, 126.9, 125.9, 114.1, 113.5, 97.6, 55.2, 28.6, 24.4, 21.9, 20.7, 16.3; IR (film): $\tilde{\upsilon}$ = 3421, 3144, 2941, 1626, 1532, 1216, 1032, 805 cm^{-1}; HRMS (ESI): *m/z*: calcd for $C_{19}H_{20}N_4O_2Na^+$: 359.1479, found: 359.1479.

1-Amino-5-(5-hydroxy-2-methylphenyl)-7,8,9,10-tetrahydropyrimido[4,5-c]isoquinolin-6(5H)-one (7)



To a solution of 1-amino-5-(5-methoxy-2-methylphenyl)-7,8,9,10-tetrahydropyrimido[4,5-c]isoquinolin-6(5H)-one (127 mg, 0.4 mmol) in CH_2Cl_2 (4 mL) was added BBr_3 (1 M, 944 µL , 0.9 mmol). The reaction was stirred at 25 °C for 16 h, and poured into MeOH. The reaction mixture was evaporated several times EtOH, triturated with EtOAc, and the precipitate was fil tered off. Recrystallization in EtOH and CH_2Cl_2 afforded the expected product as a white soli d (34 mg, 28%). mp 262-265 °C; ^{1}H NMR (400 MHz, DMSO-d_6): δ = 9.33 (s, 1H), 8.04 (s, 1 H), 7.15-7.09 (m, 3H), 6.72 (dd, J = 8.2 Hz, J = 2.5 Hz, 1H), 6.41 (d, J = 2.5 Hz, 1H), 3.07-3. 01 (m, 2H), 2.46-2.43 (m, 2H), 1.74-1.71 (m, 7H); ^{13}C NMR (125 MHz, DMSO-d_6): δ = 160. 9, 159.7, 155.9, 154.7, 153.5, 142.9, 137.4, 130.7, 126.3, 125.0, 115.4, 115.0, 97.6, 28.6, 24.4, 21.9, 20.7, 16.3; IR (film): $\tilde{\upsilon}$ = 3415, 3138, 2936, 2860, 1622, 1525, 1450, 1300, 1208, 804 c m^{-1}; HRMS (ESI): *m/z*: calcd for $C_{18}H_{18}N_4O_2Na^+$: 345.1322, found: 345.1323. HPLC purity 8 8.4%.

Protein Expression and Purification

A clone of the EphA3 kinase domain (residues: 606-947) was obtained from Prof. Sirano Dhe-Paganon's group[1] and expressed in Escherichia coli strain BL21 (DE3). Cells expressing EphA3 were induced with a 1 mM solution of isopropyl-beta-D-thiogalactopyranoside (IPTG) for 12 h at 15 °C. Cell pellets were resuspended in buffer A (50 mM Tris, pH 8.0, and 100 mM NaCl, supplemented with protease inhibitors) and lysed by sonication. After centrifugation at 15,000 rpm for 1 h, the soluble fraction of EphA3 was purified using HisTrap FF crude and HiTrap Q HP columns (GE Healthcare), followed by gel filtration chromatography (Superdex75; GE Healthcare). The appropriate fractions were combined and concentrated to ~10 mg/mL using Amicon filter devices (10 kDa as cutoff) in a storage solution (100 mM sodium chloride and 10 mM Tris-HCl pH 8.0, 5% glycerol). The resulting solution was aliquoted and stored at -80 °C for further usage.

Crystallization, Data Collection, and Structure Determination

Crystals of the EphA3 kinase domain were grown at 20 °C using the hanging drop vapor diffusion method. A 5 mM compound **7** (in 100 % DMSO) was added into the EphA3 protein to reach a final DMSO concentration of 10% (v/v) and the mixture was incubated on ice for 1 hour before crystallization. Then equal volumes of protein (with compound **7**) and reservoir solutions (0.1 M sodium cacodylate pH 6.5, 0.15 M ammonium sulfate, 22.5% PEG 3350) were mixed and crystals appeared after 1 to 2 days. The crystals were flash-frozen in liquid nitrogen without extra cryoprotectant for measurements.

Crystallography

Data sets were collected on a PILATUS 2M detector at the Swiss Light Source beamline X06DA of the Paul Scherrer Institute (Villigen, Switzerland) and indexed, integrated and scaled with the XDS[2] and CCP4[3] programs. The structures were solved by molecular replacement with PHASER[4] using the apo EphA3 kinase domain structure (PDB entry 2GSF) as a search model and refined with PHENIX[5].

Space group P1 21 1		<I/σ(I)>	24.5(4.3)
Unit cell		R merge	0.044(0.416)
a (Å)	53.07	Completeness (%)	99.6(97.8)
b (Å)	38.24	Multiplicity	6.6(6.1)
c (Å)	75.84	**Refinement**	
alpha	90.00	Resolution range (Å)	39.07-1.70
beta	101.59	R factor/R free	0.179/0.208
gamma	90.00	Mean B factors (Å^2)	26.40
Resolution range (Å)	47.29 -1.70	RMS bonds (Å)	0.0106
Unique reflections	33115(4694)	RMS angles (°)	1.510

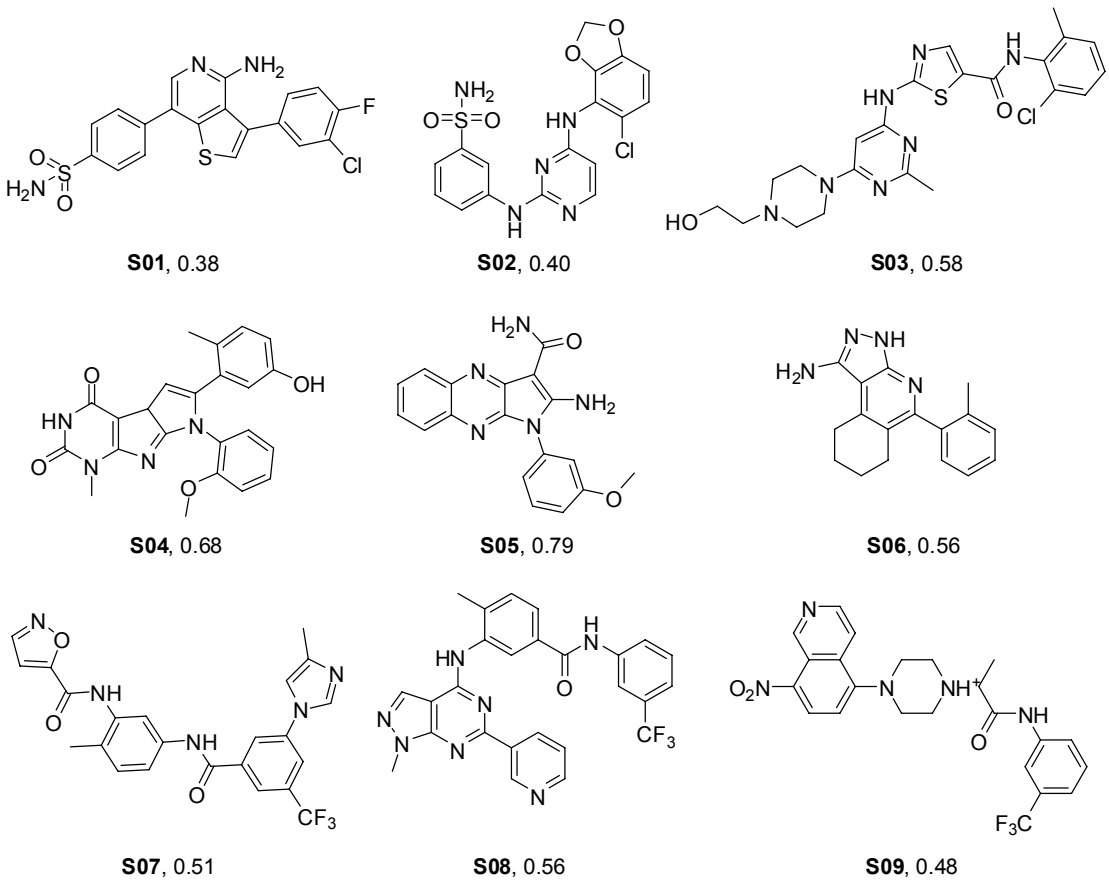


Figure S1. Previously reported Eph inhibitors. [1, 6-13] The values next to the compound number are the Tanimoto similarity index calculated by MACCS keys with compound **7**. Compound **S04**, **S05** and **S09** were discovered by our in silico approach.[9-10, 13]

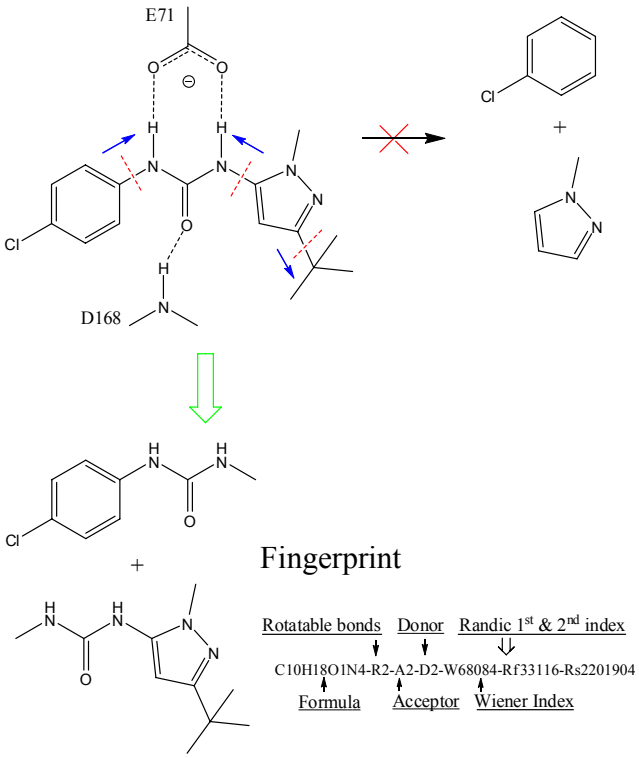


Figure S2. Decomposition of a known type II kinase inhibitor (PDB entry 1KV1) into anchor fragments. Over-decomposition could lead to very small fragments (top, right). Here, the following protocol is used to generate anchor fragments by decomposition of a compound library. Firstly, all rotatable bonds of a molecule are broken to obtain initial ring systems. Secondly, each ring system is extended at the cut until it reaches another ring or the length of the extension (number of heavy atoms) reaches three. In case the atom on the third level is not an sp^3 carbon, the elongation goes further until it reaches an sp^3 carbon or a ring. Thirdly, to keep the valence unchanged, hydrogen atoms are attached to carbon atoms while heteroatoms are capped by CH_3 groups. Lastly, to remove identical anchor fragments, a fingerprint consisting of molecular formula, number of rotatable bonds, acceptors and donors, and two 2D topological indices, i.e. modified Wiener[14] and Randic[15] index, is computed. The Wiener index is modified by taking valence and element difference into account as $W = \sum(v_i e_i \sum(d_{ij} e_j))$, wherein v_i is the valence of the i^{th} atom minus the number of hydrogen atoms attached to it, e_i is the atomic number, and d_{ij} is the distance between the i^{th} and j^{th} atom in the molecular graph. The modified Randic first and second index are ${}^1\chi = \sum p_i p_j$ and ${}^2\chi = \sum p_i p_j p_k$, respectively, with $p_i = \sum d_{ij}$.

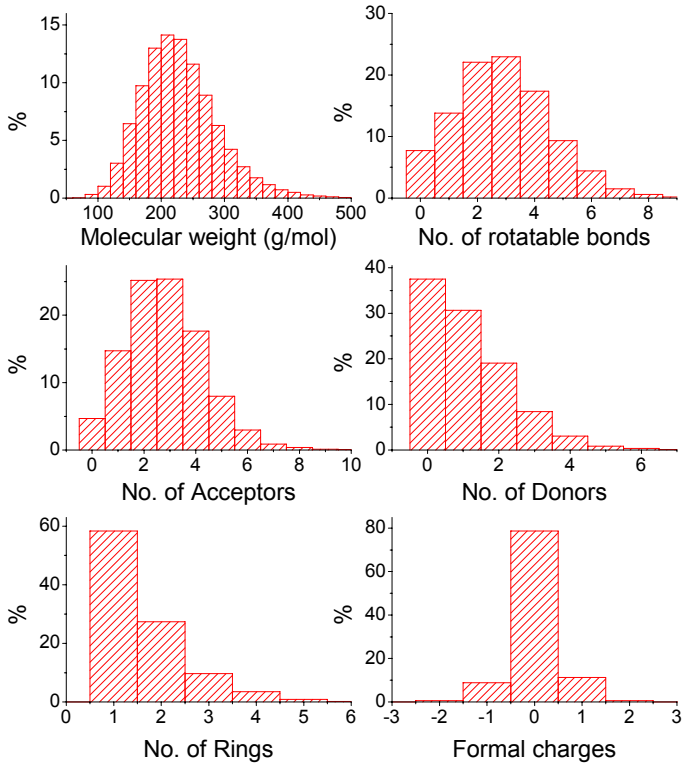


Figure S3. Distribution of physicochemical properties of the library of anchor fragments (563,774 fragments) obtained by decomposing the ZINC all-now compound library (9 million molecules) within 3 hours on a single commodity CPU by the in house developed software LIBO. The program LIBO is written in C++ language and is available upon request on the homepage of the last author.

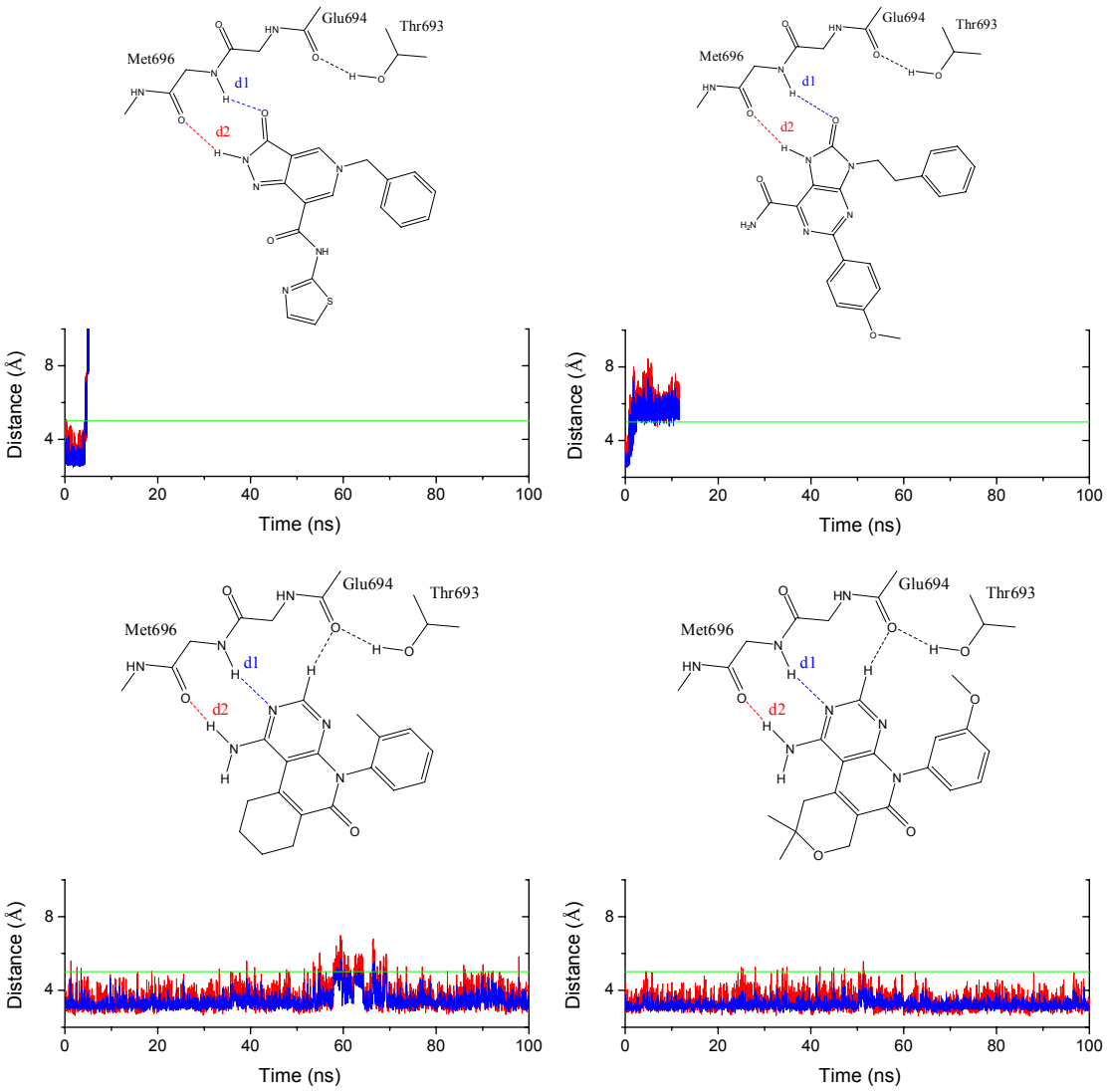


Figure S4. Screening by explicit solvent molecular dynamics. 2D representation of binding mode as well as MD time series of Hbonds of four scaffolds identified by the ALTA virtual screening approach. The two molecules in the top panels are discarded because their hydrogen bonds with the hinge region break apart within the first 10 ns.

References

1. Choi, Y.; Syeda, F.; Walker, J. R.; Finerty, P. J., Jr.; Cuerrier, D.; Wojciechowski, A.; Liu, Q.; Dhe-Paganon, S.; Gray, N. S. Discovery and structural analysis of Eph receptor tyrosine kinase inhibitors. *Bioorg Med Chem Lett* **2009,** 19, 4467-70.
2. Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* **1993,** 26, 795-800.
3. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **1994,** 50, 760-3.
4. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **2007,** 40, 658-674.
5. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** 58, 1948-54.
6. Miyazaki, Y.; Nakano, M.; Sato, H.; Truesdale, A. T.; Stuart, J. D.; Nartey, E. N.; Hightower, K. E.; Kane-Carson, L. Design and effective synthesis of novel templates, 3,7-diphenyl-4-amino-thieno and furo-[3,2-c]pyridines as protein kinase inhibitors and in vitro evaluation targeting angiogenetic kinases. *Bioorg Med Chem Lett* **2007,** 17, 250-4.
7. Bardelle, C.; Cross, D.; Davenport, S.; Kettle, J. G.; Ko, E. J.; Leach, A. G.; Mortlock, A.; Read, J.; Roberts, N. J.; Robins, P.; Williams, E. J. Inhibitors of the tyrosine kinase EphB4. Part 1: Structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg Med Chem Lett* **2008,** 18, 2776-80.
8. Sun, L.; Liang, C.; Shirazian, S.; Zhou, Y.; Miller, T.; Cui, J.; Fukuda, J. Y.; Chu, J. Y.; Nematalla, A.; Wang, X.; Chen, H.; Sistla, A.; Luu, T. C.; Tang, F.; Wei, J.; Tang, C. Discovery of 5-[5-fluoro-2-oxo-1,2-dihydroindol-(3Z)-ylidenemethyl]-2,4- dimethyl-1H-pyrrole-3-carboxylic acid (2-diethylaminoethyl)amide, a novel tyrosine kinase inhibitor targeting vascular endothelial and platelet-derived growth factor receptor tyrosine kinase. *J Med Chem* **2003,** 46, 1116-9.
9. Lafleur, K.; Huang, D.; Zhou, T.; Caflisch, A.; Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J Med Chem* **2009,** 52, 6433-46.
10. Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011,** 6, e19923.
11. van Linden, O. P.; Farenc, C.; Zoutman, W. H.; Hameetman, L.; Wijtmans, M.; Leurs, R.; Tensen, C. P.; Siegal, G.; de Esch, I. J. Fragment based lead discovery of small molecule inhibitors for the EPHA4 receptor tyrosine kinase. *Eur J Med Chem* **2012,** 47, 493-500.
12. Martiny-Baron, G.; Holzer, P.; Billy, E.; Schnell, C.; Brueggen, J.; Ferretti, M.; Schmiedeberg, N.; Wood, J. M.; Furet, P.; Imbach, P. The small molecule specific EphB4 kinase inhibitor NVP-BHG712 inhibits VEGF driven angiogenesis. *Angiogenesis* **2010,** 13, 259-67.
13. Zhao, H.; Huang, D.; Caflisch, A. Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. *ChemMedChem* **2012**.
14. Wiener, H. Structural determination of paraffin boiling points. *J Am Chem Soc* **1947,** 69, 17-20.
15. Randic, M. Characterization of molecular branching. *J Am Chem Soc* **1975,** 97, 6609-6615.

Chapter 5

Mr. Dock: Molecular Recognition in Flexible Docking of Small Molecules to Proteins

Mr. Dock: Molecular Recognition in Flexible Docking of Small Molecules to Proteins

Hongtao Zhao and Amedeo Caflisch*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abstract: Drug discovery has seen the increasingly importance of structure-based virtual screening, in which molecular docking plays a key role. Other than developing a sophisticated physics-based scoring function to improve accuracy in prediction of binding poses, here we present a new flexible docking tool (called Mr. Dock) by taking advantage of Molecular Recognition in flexible Docking of small molecules to proteins, namely shape complementarity and hydrogen bonding. Mr. Dock shows a success rate of 94.1% on the Astex diversity set and of 97% on a new test set consisting of 100 kinase complexes. Prospectively, in a recent high-throughput virtual screening campaign against zeta-chain associated protein 70 (ZAP70), application of Mr. Dock leads to the discovery of several novel classes of inhibitors with IC_{50} in the low micromolar range, indicating its promising utility in structure-based virtual screening.

Introduction

Molecular docking, as computational simulation of a candidate ligand binding to a receptor, is used to predict the binding poses of small-molecule drug candidates into their protein targets of pharmaceutical relevance. It has emerged as an efficient tool in drug discovery, from hit identification to lead optimization and beyond.[1,2] Flexible docking, in most cases so far, is limited to a chosen set of covalent bonds that can freely rotate in search of a native pose, while keeping the receptor, bond lengths and angles fixed.[3-7] To allow for more degrees of freedom e.g. flexible receptor[8] is not only computationally very expensive, but also could lead to dramatically increased false positives.

Molecular docking consists of essentially a scoring function and a random search algorithm. The scoring function, used to prioritize the poses in the process of random search, is core of molecular docking. It is typically in a form of energy function that aims to predict the binding free energy of a small molecule to the receptor. In the framework of empiric force fields, the binding free energy consists of van der Waals interaction energy, Coulombic interaction energy in vacuo, desolvation energy, conformational strain of small molecules, entropic loss of both small molecules and receptors, entropic gain by releasing constrained water molecules into the bulk, and so on. In order to improve the accuracy in prediction of the binding poses and also quite often the binding free energies, scoring functions in docking tools[3-6] are thus trying to account for all the factors mentioned above, especially the mysterious role of water molecules[5-6] in a physics-based approach.

The random search algorithm in early history is simulated annealing approach,[3] with the risk of being trapped in local minima. Nowadays, docking tools[4-7] typically adopt the evolution approach with a local optimization step during each generation. The random search algorithm has been extensively explored in computer science and protocols of optimized search can be easily found elsewhere. Finally, programming embodies a docking tool. The embodiment depends largely on the coding techniques, so that a docking tool should always be evaluated based on its performance in the field, instead of referring to its scoring function only.

Recently, Trott and Olson presented a docking tool called AutoDock Vina,[7] based on a scoring function as more “machine learning” than directly physics-based in its nature. AutoDock Vina[7] outstandingly outperforms the physics-based AutoDock 4[4], indicating “superficially physics-based scoring functions do not necessarily perform better”.[7] From the perspective of molecular recognition, molecular docking is quite intuitive that small molecules bind to the receptors firstly in a shape complementary way and then are fixed by anchors of hydrogen bonds.

Here, we report on a user-friendly and accurate docking tool (called Mr. Dock) that takes advantage of **M**olecular **R**ecognition in flexible **Dock**ing. Mr. Dock was validated on the Astex diversity set[9] with a success rate of 94.1% using biased geometry, and on the Kinase 100 set[10] with a rate of 97% using unbiased geometry. Prospectively, in a recent high-throughput virtual screening campaign against zeta-chain associated protein 70 (ZAP70), which plays a crucial role in B- and T-cell activation,[11-12] application of Mr. Dock leads to the discovery of several novel chemotypes with IC_{50} in the low micromolar range, indicating its promising utility in structure-based virtual screening.

Methods

Scoring Function. The following scoring function is implemented in Mr. Dock:

$$Score = 0.18Vdw + 0.6Hb + 0.018Coul + 0.36N_{\text{inter-clash}} + 0.24N_{\text{intra-clash}} \quad (1)$$

$$Hb = w_d * w_a * (r - 2.8) \quad (2)$$

wherein, *Vdw* is the van der Waals interaction energy calculated by the corresponding equation in the CHARMM force field[13]; *Hb* is the hydrogen bonding energy calculated by equation **2**; *Coul* is the Coulombic interaction energy by a distance dependent dielectric of *r*, and the first three terms were calculated among non inter-clash ligand atoms. Inter-clash is defined if interaction energy of the ligand atom with the protein is greater than 0.36 kcal/mol. $N_{\text{inter-clash}}$ is the number of ligand atoms having clash with the protein; $N_{\text{intra-clash}}$ is the number of pairs of ligand heavy atoms having intra-molecular clash, which is determined by a minimum pairwise distance of 3 Å or the shorter distance in the initial ligand conformation to allow for intra-molecular hydrogen bonds. w_d and w_a is the hydrogen bonding weight of donors and acceptors,[10] *r* in equation **2** is the distance between the donor hydrogen atom and the acceptor, and *HB* equals 0 if $r < 1.6$ Å or $r > 2.8$ Å.

The van der Waals coefficient is directly derived from the previous scoring function incorporating hydrogen bonding penalty[10], and the hydrogen bonding coefficient is taken from estimate of the net gain by formation of one hydrogen bond, around 0.6 kcal/mol. It has been observed that small coefficients of clash facilitate the simulated annealing search but smaller values deteriorate the poses, thus the coefficients of clash are determined according to the search performance. Partial charges were assigned only to donor hydrogen atoms and acceptors, with the aim to optimize the geometry of local polar interactions. The van der Waals attraction of carboxylic oxygen atom is set to zero in order to prevent from burial deeply inside the protein.

To speed up, energy of each ligand atom type (derived from the CHARMM general force field[14] by merging similar ones, Table 1) in the binding site was calculated and stored in a

grid-based manner. The grid maps are calculated internally like in AutoDock Vina[7], with a fine grid space of 0.2 Å. The energy of each atom type is then retrieved to get the binding energy during conformational search.

Table 1. Defined ligand atom types.

Element	H	C	N	O & F	P & I	Br & S	Cl
ε_{min} (kcal/mol)	-0.046	-0.062	-0.20	-0.12	-0.55	-0.42	-0.32
r_{min} (Å)	0.55	2.00	1.60	1.50	2.18	2.07	1.93

Hybridized Search Algorithm. The genetic algorithm is adopted, with the first generation obtained from simulated annealing search. At the beginning of each simulated annealing search, the input ligand conformation, including its position, orientation and torsions, is randomly changed, so that each search starts with a different ligand conformation.

Implementation. Mr. Dock was coded in C++ language, with a simplified input interface: size of the binding site, number of runs, root mean square deviation (RMSD) cutoff for clustering of poses, and a ligand list for continuous docking without re-calculating the grid maps. Unlike AutoDock[7-8], Mr. Dock directly reads small molecules in a TRIPOS mol2 format, thus eliminating the step of preprocessing small molecules by a different tool. However, the protein still needs to be prepared by a third-party tool to add hydrogen atoms, and assign both partial charges and CHARMM/m atom types. In all, the time spent on developing Mr. Dock together with the following validation on a set of about 200 protein-ligand complexes was within one month.

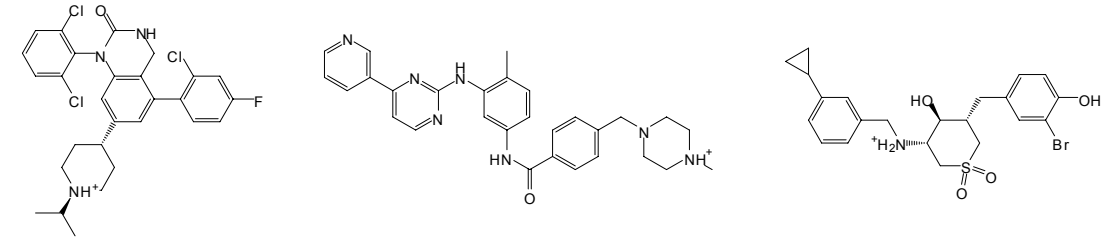
Docking Setup

Training Set. The training set includes 10 complexes including kinases, β-secretase, HIV-protease, influenza virus neuraminidase, dihydrofolate reductase and thrombin, which are of interest to the majority of medicinal chemists. The 2D structures of the complexed inhibitors are shown in Figure 1. Flexible docking of such inhibitors is generally regarded as tough cases,[4] given the number of active rotatable bonds.

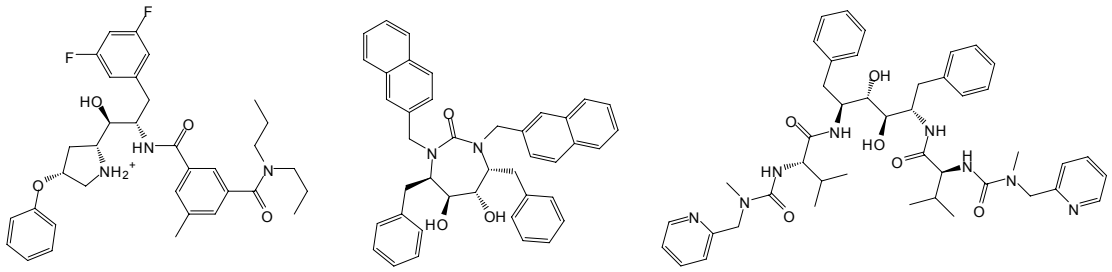
Test Set. The test set includes the Astex diversity set[9] and a new test set consisting of kinases only[10]. The Astex diversity consists of 85 proteins of drug or agrochemical targets, with almost no global sequence similarity.[9] Since our primary interest is on protein kinases that are involved in one third of current drug discovery programs[15], we further include 100 kinases (Figure 2), previously used to analyze hydrogen bonding penalty upon ligand binding[10].

Protein Preparation. All water molecules were removed and metal ions coordinating with ligand were kept. Hydrogen atoms were added according to the protonation states of chemical groups at pH 7, MPEOE partial charges[16] and CHARMm atom types[17] were assigned by the program WITNOTP. Alternatively, the protein structure can be prepared by the program VMD[18] with both partial charges and CHARMM atom types[13] assigned.

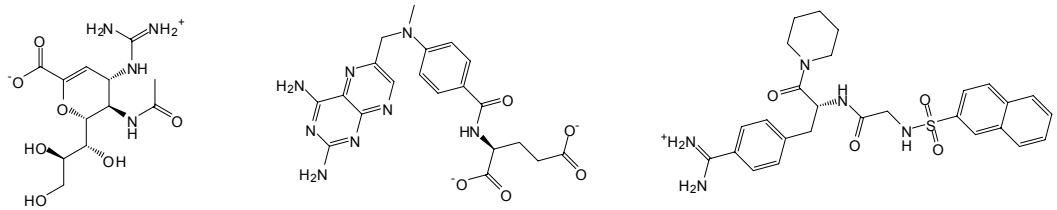
Docking. The binding site was determined as within 4 Å from any atom of the complexed ligand. The number of runs was set as 20 and RMSD cutoff as 1 Å.



1 3GC7 **2** 1OPJ **3** 3PI5



4 2QMF **5** 1HVR **6** 1HVS



7 2QWE **8** 4DFR **9** 1ETS

Figure 1. 2D representations of small molecules in the training set including two kinase inhibitors (PDB code 3GC7 and 1OPJ), two *β*-secretase inhibitors (3PI5 and 2QMF), two HIV-protease inhibitors (1HVR and 1HVS), one influenza virus neuraminidase inhibitor (2QWE), one dihydrofolate reductase inhibitor (4DFR) and one thrombin inhibitor (1ETS).

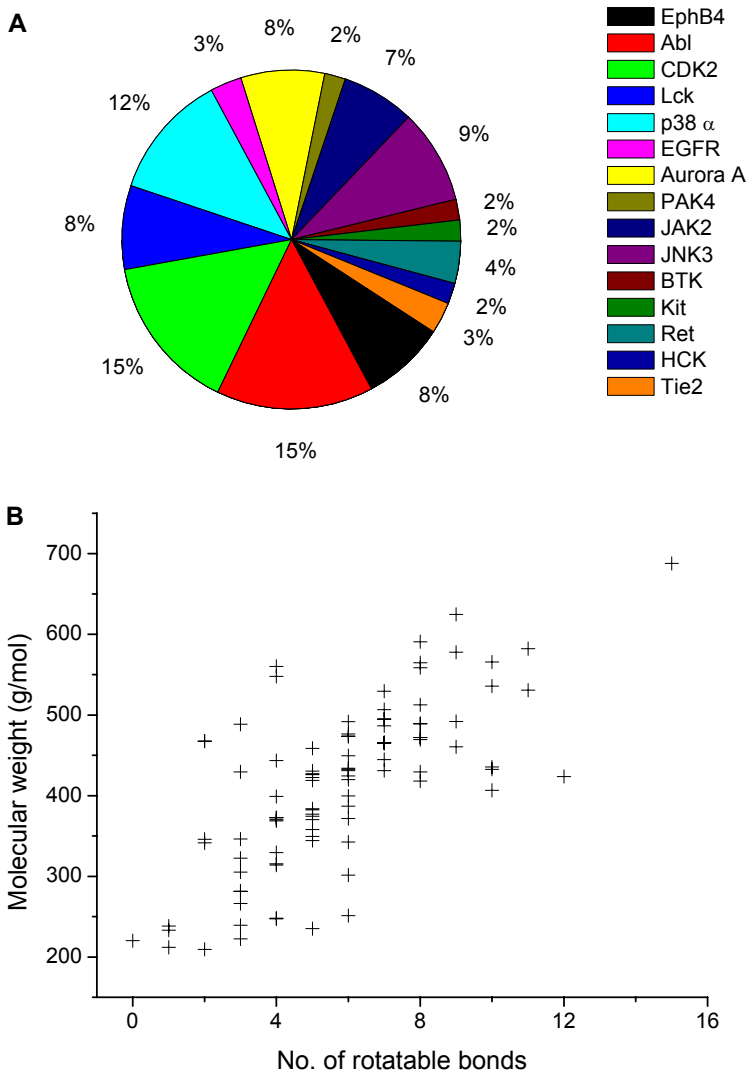


Figure 2. Profile of the Kinase 100 set.

Results and Discussions

Shape Complimentarity. Shape complimentarity between ligands and receptors is an essential aspect of molecular recognition,[19] and there are numerous cases where natural substrates bind to the specific shape of the binding site.[20-21] At the very beginning, the scoring function consists of only van der Waals interaction energy. Surprisingly, we found that native poses in the crystal structures for the training set still can be reproduced although with low convergence i.e. that the probability of reproducing the native pose from multiple runs is low. It gives moderate convergence for two kinase inhibitors (3GC7 and 1OPJ) and one HIV protease inhibitor (1HVR). The complexed inhibitor in PDB entry 3GC7 is quite rigid in a unique shape, thus binding to the ATP site of the kinase in a specific way although the binding site itself is an open cleft without imposing steric constraints on the ligand shape. While for the inhibitor complexed in 1OPJ (7 rotatable bonds) or 1HVR (10 rotatable bonds),

the inhibitor itself is quite flexible, but the specific shape of the binding site restricts the conformational space that the ligand can adopt upon binding. In these three cases, the search guided by van der Waals interaction only is able to locate the native pose with a moderate convergence, due to shape complimentarity that requires no clash between ligands and receptors.

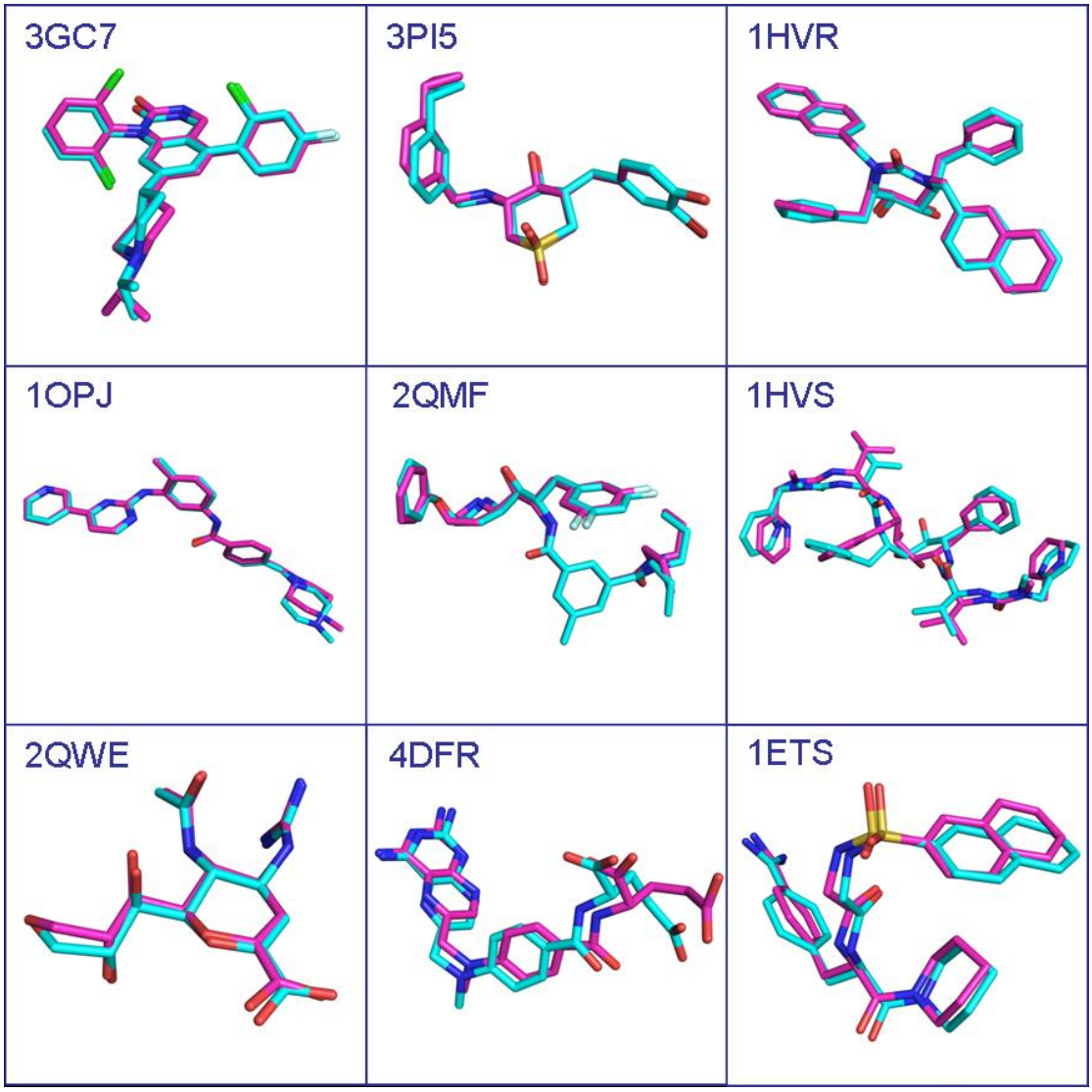


Figure 3. Best predicted pose (magenta) superimposed on the native pose (cyan) in the training set.

Hydrogen Bonding. The dissociation energy of a hydrogen bond covers more than two orders of magnitude from about 0.2 to 40 kcal/mol,[22] and moderate hydrogen bonds in the biological system resembling those between water molecules or in carbohydrates are associated with energies in the range from 4 to 15 kcal/mol.[22-23] Therefore, hydrogen bonding is another important aspect of molecular recognition. In the scoring function of AutoDock,[3] an energy function in a form similar to the van der Waals equation is used for hydrogen

bonds. Later on in AutoDock Vina,[7] a linear distance dependent function is adopted. Here, we keep the linear function form, but with introducing hydrogen bonding weights based on the consideration that strength of a hydrogen bond relies on the specific donor and acceptor pair. The concept of hydrogen bonding weights has been practiced in our early work,[10] and in the current application we expanded the Hbond atom types to better distinguish donors or acceptors in different chemical environments (data not shown). The values of the hydrogen bonding weights were derived from our chemical intuitions and experiences in structure-based virtual screening,[10, 24-25] and further optimized in a trial and error procedure with the aim to reproduce the native poses. The metal ions are treated as hydrogen bonding donors like in X-score.[26]

Table 2. Docking performance in the training set.

Protein	PDB code	Rotatable bonds	RMSD (Å)[a,b]	$f_{RMSD \leq 2.0Å}$[a,c] (%)	Time per run (minutes)
Kinase	3GC7	4	0.63	100	0.40
	1OPJ	7	0.38	85	0.57
β-secretase	3PI5	8	0.45	25	0.55
	2QMF	14	0.52	30	0.70
HIV protease	1HVR	10	0.37	95	0.63
	1HVS	21	1.19	5	1.33
Influenza virus neuraminidase	2QWE	9	0.36	85	0.40
Dihydrofolate reductase	4DFR	11	1.35	35	0.47
Thrombin	1ETS	9	0.58	10	0.58

[a]RMSD calculated by ligand heavy atoms with respect to its native pose. [b]by the best pose. [c]Percentage of poses with RMSD within 2 Å. The RMSD cutoff of 2 Å is often used as a criterion of the correct bound structure prediction.

During simulated annealing search, small clash coefficients allow to quickly locate non-clash conformations by lowering the barriers; while hydrogen bonding interaction traps the ligand in a conformation in favor of good hydrogen bonding pattern. Hydrogen bonds has long been recognized for its role in providing binding specificity, and it is indeed more efficient to incorporate hydrogen bonding interaction by directing the search towards the native pose. As a result, it leads to a better convergence with significantly improved accuracy (Table 2 and Figure 3), even lack of a local optimization step in the evolution search. Without hydrogen bonding weights, the search could be trapped into unrealistic conformations which

simply favor more hydrogen bonds even rather weak, decreasing search efficiency and sometimes even failing to reproduce the native pose.

Table 3. Docking performance on Astex diversity set and Kinase 100 set.[a]

	$f_{RMSD\leq2.0Å}$[b] (%)	$f_{RMSD\leq2.0Å}$[c] (%)	Time per run[d] (minutes)
Astex diversity set[9]	94.1 ± 2.0[e]	82.4 ± 0.0[e]	0.28
Kinase 100 set[10]	97	91	0.40

[a]For Astex diversity set, the geometry of native poses was used as input; while for Kinase 100 set the unbiased geometry was used by minimizing the native poses with the CHARMm force field in the absence of the proteins. In both cases, the native poses were reoriented manually to remove the memory of binding. [b]by the best pose among 20 runs. [c]by the first-ranked pose among 20 runs. [d]excluding the time of processing proteins which can be calculated just once in a high-throughput screening. [e]standard deviation based on three replicates with each consisting of 20 runs.

Validation on Test Sets. Compared with the success rate of 86.5% by GOLD on the Astex diversity set,[9] Mr. Dock gives a rate of 94.1% by the best pose and 82.4% by the first-ranked pose (Table 3 and Figure 4).

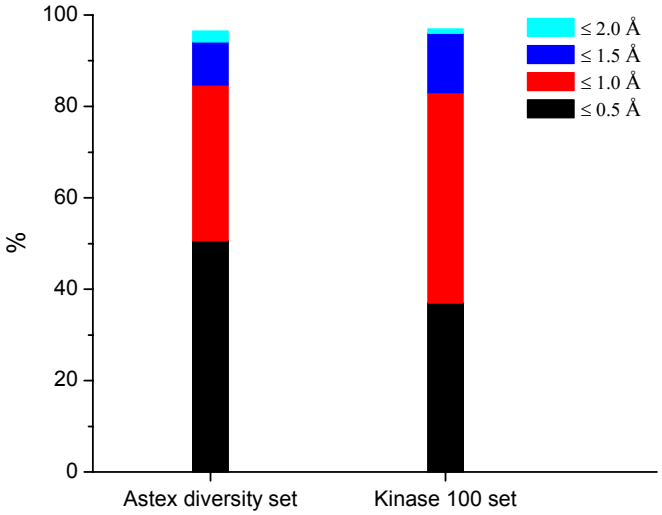


Figure 4. Distribution of RMSD of the best pose among 20 runs relative to the native pose in the crystal structure.

In re-docking of the Astex diversity set, both the position and orientation of the native poses were changed randomly in order to remove the memory of binding, however, the geometry was kept unchanged. It has been pointed out that conformation of a bound ligand could be different from its unbound state and even small changes in bond angles could lead to failure of re-docking.[27] Keeping this in mind, we minimized the native poses in the absence of proteins to get unbiased geometry for the Kinase 100 set before re-docking. Notably, Mr.

Dock gives a success rate of 97% by the best pose and 91% by the first-ranked pose (Table 3 and Figure 4).

Not surprisingly, Mr. Dock performs better on the Kinase 100 set, and especially on the type II kinase inhibitors even they typically possess more rotatable bonds than type I inhibitors. Kinase inhibitors are in general hinge binders which form two or three hydrogen bonds with hinge of the ATP site. With the flip of the activation loop when kinase is in its inactive state, the type II inhibitors take advantage of the exposed allosteric site and form two or three additional hydrogen bonds with the protein.[15] Hydrogen bonding together with shape complementarity then can well position the kinase inhibitors into the binding site. Indeed, use of hydrogen bonds has become one of our routine procedures in discovery of novel kinase inhibitors.[10, 24-25]

The failure on a few cases from the Astex diversity set (e.g. 1HVY and 1XM6) originates from the absence of a robust desolvation term in the scoring function. In the absence of crystal waters, ligands tend to be in a close contact with the proteins to gain favorable van der Waals interaction, which should be prevented otherwise if a desolvation penalty is present. It has been shown that presence of crystal waters improves re-docking accuracy.[28] Aside from the mysterious magic of water molecules, it is not surprising at all since presence of crystal waters significantly reduces the conformational space of a bound ligand by imposing steric constraints, turning the binding site into a template of the ligand's shape.[9]

The average computational time per run on a commodity computer (a single Intel 2.8 GHz CPU) is 0.28 minutes for the Astex diversity set, and 0.40 minutes for the Kinase 100 set. The memory usage varies from about 40 to 90 MB, depending on the size of the binding site.

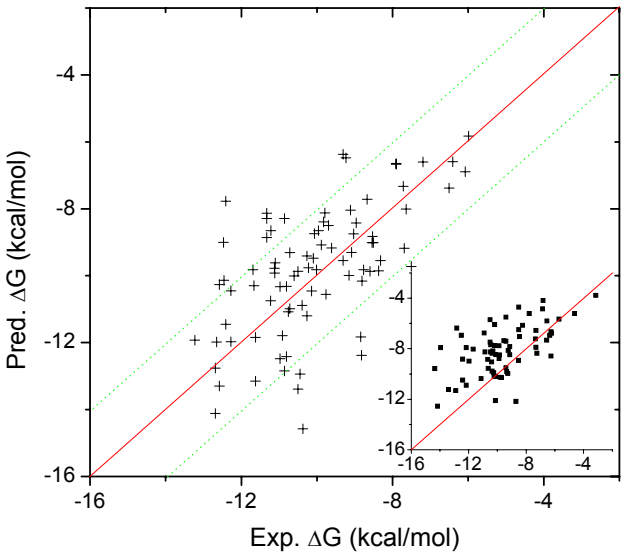


Figure 5. Correlation between the predicted and experimental binding affinities for the Kinase 100 set. Inset: Astex diversity set.

Scoring Function. The scoring function **1** was empirically derived based on our experiences in discovery of kinase inhibitors,[10, 24-25] without further fitting. The kinase binding site is predominantly hydrophobic,[29] and the electrostatic contribution as a sum of Coulombic interaction in vacuo and desolvation is marginal.[10, 29] In Mr. Dock, partial charges were assigned only to donor hydrogen and acceptor atoms with the purpose of optimizing local polar interactions, and the electrostatic coefficient is only 10% van der Waals coefficient. The electrostatic contribution in Mr. Dock is thus consistent with our observation on kinase inhibitors binding. Notably, Mr. Dock achieves a root mean square error of 1.83 kcal/mol for the Kinase 100 set (Figure 5). In contrast, Mr. Dock tends to underestimate the binding affinity in the Astex diversity set in the case where hydrophilic interaction is dominant.

Unlike other docking tools[3-7], prediction of binding affinities is no longer the pursuit of Mr. Dock. In order to improve computational efficiency, a docking tool often introduces many approximations and sometimes "machine-learning" elements into the scoring function.[7] We suggest that a more sophisticated energy function should be used to predict the binding affinity upon energy minimization following docking, and use a docking tool only to generate binding poses, as shown in our previous practices.[10, 24-25] In this sense, we expect only moderate convergence of docking, so that a diversity set of poses can be available for subsequent evaluation by a more accurate energy function.

Prospective Application of Mr. Dock. Limited prospective applications do not necessarily indicate the validity of a tool or method; however, it is our only interest to develop Mr. Dock for its application in the field. Firstly, rational binding poses of three series of kinase inhibitors[10, 24-25] including one type II inhibitor[24] (the predicted binding poses of two nanomolar inhibitors[10, 25] have been confirmed by X-ray), previously discovered by AutoDock 4[4], were successfully reproduced by Mr. Dock. Secondly, we applied Mr. Dock in a recent high-throughput virtual screening of novel inhibitors for zeta-chain associated protein 70 (ZAP70), for which a cell-permeable and highly specific inhibitor has not yet been reported.[11-12] Several novel chemotypes with IC_{50} in the low micromolar range were discovered (H. Zhao, in preparation), indicating its promising utility in structure-based virtual screening.

Conclusions

In summary, we present a flexible docking tool that takes advantage of molecular recognition in docking of small molecules to proteins, with comparable or slightly better performance. However, hydrogen bonding weights, partial charges, and details in the search

algorithm are not in the scope of discussion for the following reasons: 1) It follows our simple chemical and physical intuitions, as exemplified previously.[10] 2) In the optimization of docking performance, we found that generally a success rate of more than 80% can be achieved as long as hydrogen bonding interaction was included, indicating a wide range for choice of different combinations of hydrogen bonding weights. We expect that this information would be enough for developers. 3) We want to keep it in a concise and clear way with emphasis on the docking performance, which is of sole interest to the majority of practitioners in the field of medicinal chemistry.

Acknowledgment

We thank Dr. Ting Zhou and Dr. Danzhi Huang for interesting discussions. We are grateful to Armin Widmer (Novartis Basel) for the continuous support with the program WITNOTP.

References

1. Bajorath, J., Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* **2002,** 1, (11), 882-94.
2. Langer, T.; Hoffmann, R. D., Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Des* **2001,** 7, (7), 509-27.
3. Goodsell, D. S.; Olson, A. J., Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990,** 8, (3), 195-202.
4. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* **1998,** 19, (14), 1639-1662.
5. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** 47, (7), 1739-49.
6. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006,** 49, (21), 6177-96.
7. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** 31, (2), 455-61.
8. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **2009,** 30, (16), 2785-91.
9. Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W., Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem* **2007,** 50, (4), 726-41.
10. Zhao, H.; Huang, D., Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011,** 6, (6), e19923.
11. Wang, H.; Kadlecek, T. A.; Au-Yeung, B. B.; Goodfellow, H. E.; Hsu, L. Y.; Freedman, T. S.; Weiss, A., ZAP-70: an essential kinase in T-cell signaling. *Cold Spring Harb Perspect Biol* **2010,** 2, (5), a002279.

12. Au-Yeung, B. B.; Levin, S. E.; Zhang, C.; Hsu, L. Y.; Cheng, D. A.; Killeen, N.; Shokat, K. M.; Weiss, A., A genetically selective inhibitor demonstrates a function for the kinase Zap70 in regulatory T cells independent of its catalytic activity. *Nat Immunol* **2010,** 11, (12), 1085-92.
13. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998,** 102, (18), 3586-3616.
14. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010,** 31, (4), 671-90.
15. Liu, Y.; Gray, N. S., Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol* **2006,** 2, (7), 358-64.
16. No, K. T.; Grant, J. A.; Scheraga, H. A., Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method .1. Application to Neutral Molecules as Models for Polypeptides. *J Phys Chem* **1990,** 94, (11), 4732-4739.
17. Momany, F. A.; Rone, R., Validation of the General-Purpose Quanta(R)3.2/Charmm(R) Force-Field. *J Comput Chem* **1992,** 13, (7), 888-900.
18. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** 14, (1), 33-8, 27-8.
19. Nicholls, A.; McGaughey, G. B.; Sheridan, R. P.; Good, A. C.; Warren, G.; Mathieu, M.; Muchmore, S. W.; Brown, S. P.; Grant, J. A.; Haigh, J. A.; Nevins, N.; Jain, A. N.; Kelley, B., Molecular shape and medicinal chemistry: a perspective. *J Med Chem* **2010,** 53, (10), 3862-86.
20. Mattevi, A.; Fraaije, M. W.; Mozzarelli, A.; Olivi, L.; Coda, A.; van Berkel, W. J., Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. *Structure* **1997,** 5, (7), 907-20.
21. Borngraber, S.; Browner, M.; Gillmor, S.; Gerth, C.; Anton, M.; Fletterick, R.; Kuhn, H., Shape and specificity in mammalian 15-lipoxygenase active site. The functional interplay of sequence determinants for the reaction specificity. *J Biol Chem* **1999,** 274, (52), 37345-50.
22. Steiner, T., The hydrogen bond in the solid state. *Angew Chem Int Ed Engl* **2002,** 41, (1), 49-76.
23. Jeffrey, G. A., An Introduction to Hydrogen Bonding. In Oxford University Press: Oxford, 1997.
24. Zhao, H.; Huang, D.; Caflisch, A., Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. *ChemMedChem* **2012**.
25. Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A., Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *Acs Med Chem Lett* **2012**.
26. Wang, R.; Lai, L.; Wang, S., Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **2002,** 16, (1), 11-26.
27. Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A., Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J Comput Chem* **2004,** 25, (3), 412-22.
28. Thilagavathi, R.; Mancera, R. L., Ligand-protein cross-docking with water molecules. *J Chem Inf Model* **2010,** 50, (3), 415-21.
29. Kolb, P.; Huang, D.; Dey, F.; Caflisch, A., Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem* **2008,** 51, (5), 1179-88.

Chapter 6

Discovery of Novel Classes of ZAP70 Inhibitors by Docking into a Protein Conformation Generated by Molecular Dynamics

Discovery of Novel Classes of ZAP70 Inhibitors by Docking into a Protein Conformation Generated by Molecular Dynamics

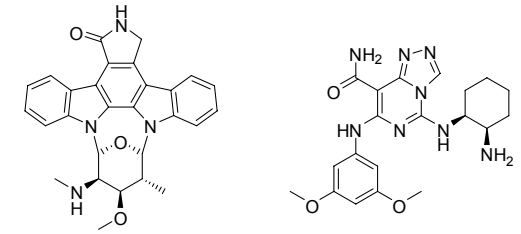
Hongtao Zhao and Amedeo Caflisch*

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

ABSTRACT: We have discovered six novel classes of ZAP70 inhibitors by docking into a protein conformation generated by molecular dynamics. Molecular dynamics simulation of ZAP70 with a potent JAK2 inhibitor was carried out to induce a fit of the gatekeeper residue Met414 and the back pocket. Two compounds with IC_{50} of about 10 μM were predicted to occupy the back pocket, and the well-established optimization scheme on the moieties in the back pocket suggests that potent ZAP70 inhibitors might be obtained by the addition of a hydroxyl group. The finding of a novel 110 nM JAK2 inhibitor on the other hand emphasizes the importance of including kinases sensitive to chemical inhibition in the screening panel.

KEYWORDS: *ZAP70, virtual screening, molecular dynamics, autoimmune disease*

Zeta-associated protein 70 kDa (ZAP70), a member of the Syk family and non-receptor protein tyrosine kinases, mainly expressed on T cells and natural killer cells, plays a crucial role in immune response and is implicated in inflammatory and autoimmune disease.[1-2] ZAP70 is an attractive therapeutic target for various allergic and autoimmune disorders as well as organ transplant medication. Although four series of small-molecule ZAP70 inhibitors have been discovered (Figure 1),[3-7] a cell-permeable and highly specific ZAP70 inhibitor has not yet been reported.[2] Most kinase inhibitors promiscuously inhibit multiple kinases, but a recent comprehensive selectivity profiling of 178 commercial available kinase inhibitors on a panel of 300 recombinant protein kinases revealed that ZAP70 is less druggable by chemical inhibition,[8] which is further confirmed by a second profiling of 72 inhibitors with 442 kinases.[9]



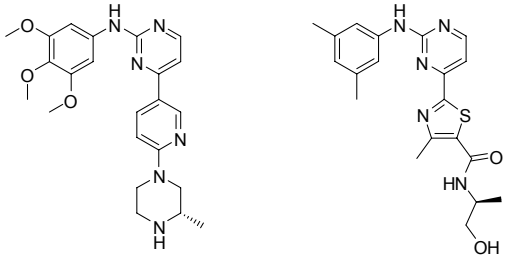


Figure 1. Four series of nanomolar ZAP70 inhibitors.

Previously we have discovered three novel classes of tyrosine kinase inhibitors,[10-12] which bind into the ATP site occupying the hydrophobic back pocket (PDB code 4G2F[12] and 4GK2). Optimization of the moieties in the back pocket quickly leads to potent and selective tyrosine kinase inhibitors[12-13] although the two initial hits of different chemotype are weak binders (Figure 2).[10, 12] More examples can be found elsewhere that targeting the back pocket is a rather efficient way to gain both potency and selectivity in kinase drug design.[14]

The access of the back pocket is modulated by a so called gatekeeper residue. ZAP70 has a bulky gatekeeper residue Met414, which blocks the entry of the back pocket and thus makes it inaccessible by small molecules as shown in two available crystal structures so far (PDB code 1U59[4] and 2OZO[15]). However, analysis of complex structures of kinases with a bulky Met gatekeeper residue in the Protein Data Bank reveals that the back pocket of some kinases such as JAK2[16] and JNK3[17] still can be accessed by few potent small-molecule inhibitors.

Although belonging to a different subfamily and sharing low overall sequence identity with ZAP70, JAK2 (PDB code 3KCK[16]) shows a high geometric similarity in the ATP site with ZAP70 (PDB code 1U59[4]). Manual placement of the complexed JAK2 inhibitor[16] (Figure 3) into the ZAP70 ATP site by aligning the two crystal structures encounters slight van der Waals clashes. Following the protocol previously presented,[18] molecular dynamics simulation of ZAP70 complexed with this putative ligand was carried out to induce a fit of the gatekeeper residue Met414 and the back pocket, so that the slight clashes can be accommodated by reorganization of the protein to allow exploration of the back pocket in virtual screening. One single snapshot called as MD-IF (molecular dynamics induced fit) structure was then selected for subsequent high throughput virtual screening, based on evaluation of hydrogen bonds formed between the putative ligand and ZAP70.

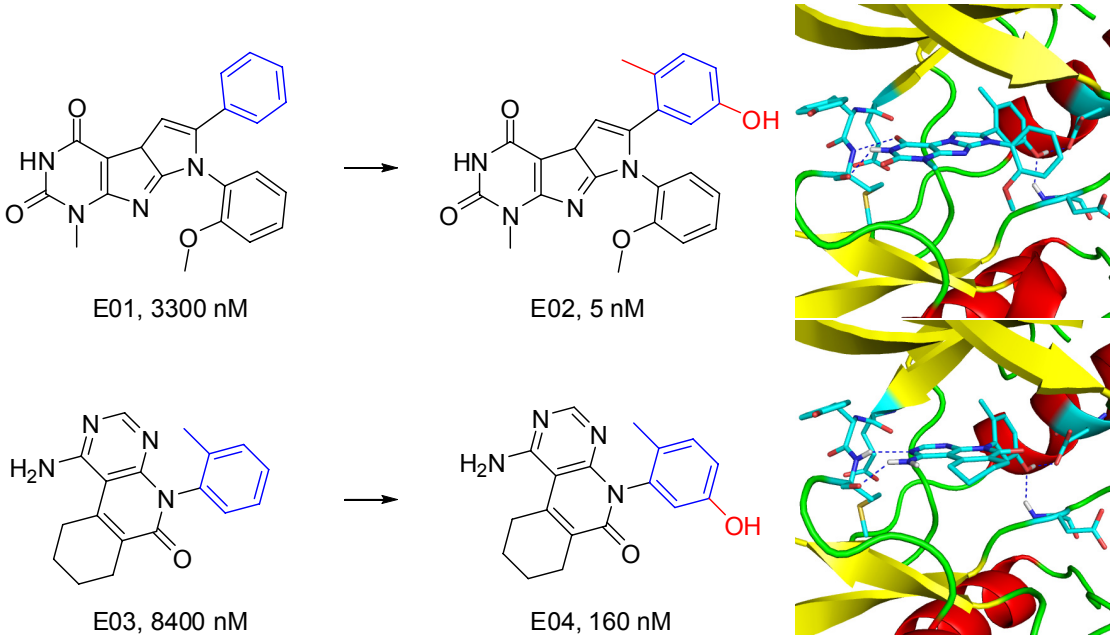


Figure 2. Hit to lead by optimization of the moieties (colored in blue) in the back pocket, and crystal structures of E02 (4GK2) and E04 (4G2F) bound to the kinase domain of EphA3. Dashed lines represent intermolecular hydrogen bonds.

The ZINC-all now library[19] (version of 2011) was first filtered to generate a kinase focused library by physicochemical properties: 1) Mw less than 380 g/mol; 2) number of rotatable bonds less than 8; 3) at least one donor and two acceptors; 4) number of rings from 3 to 5. The initial library of nearly 9 million compounds was reduced to a focused library of about 750,000, which was then docked into the MD-IF structure by a recently developed fast docking tool Mr. Dock (Mr. Dock: Molecular Recognition in Flexible Docking of Small Molecules into Proteins, Zhao and Caflisch, unpublished). Mr. Dock adopts a combination of simulated annealing and genetic algorithm optimization of position, orientation, and rotatable bonds of the ligand, with a successful rate of 94.1% on the Astex diversity set[20] and of 97%

on the 100 kinase complexes.[11] The docking poses were further minimized in the rigid protein by CHARMM[21-22] with the CHARMm22 force field.[23] The poses were then subject to three filters: van der Waals efficiency of -0.1 kcal/g,[18] hydrogen bonding penalty[11] of 2, and hydrogen bonds to the hinge, to gain computational efficiency by removing unrealistic poses before rescoring by a sophisticated energy function. Afterwards, the passed compounds were ranked according to a previously reported scoring function,[11] which has a proved track record in discovery of novel classes of kinase inhibitors with activity ranging from low micromolar to nanomolar.[11-12, 18] Finally, 32 compounds were selected and purchased for bioassay. The process of virtual screening is briefly illustrated in Figure 3.

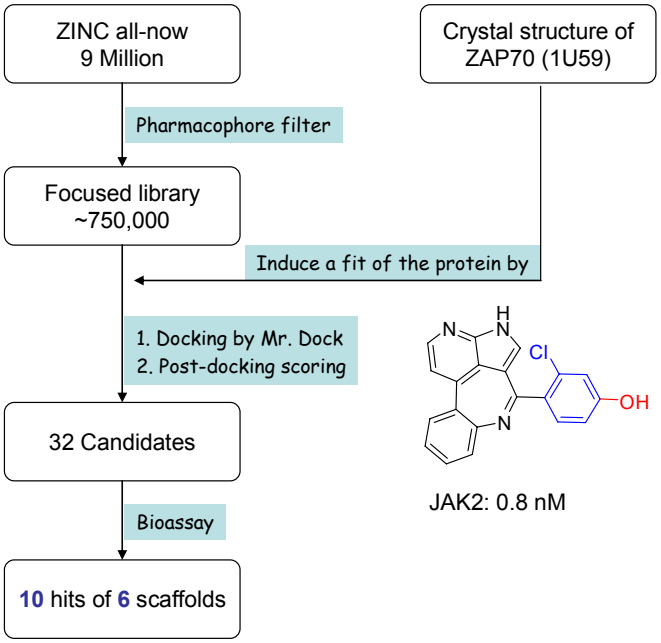


Figure 3. Flow chart of virtual screening.

The IC_{50} measurements on full-length ZAP70 were carried out at Reaction Biology Corp. with radioactive ATP at 1 μM concentration. Notably, 10 compounds of 6 novel scaffolds showed low micromolar inhibitory activity against ZAP70, with the most active being 8.5 μM. Figure 4 shows the representative compound of each scaffold with IC_{50} values listed in Table 1. All the six compounds are predicted to form at least two hydrogen bonds with Ala417 of the hinge, a segment linking the N- and C-lobe of the kinase domain (Figure 5 and 6). Compounds **01**, **03** and **04** form an additional hydrogen bond with Glu415 of the hinge. Such hydrogen bonding pattern is a typical feature of hinge binders, mimicking ATP binding. Compounds **01** to **04** do not occupy the hydrophobic back pocket, and the same binding modes can be reproduced by docking into the original crystal structure (1U59). In contrast, compounds **05** and **06** utilize the back pocket, and thus cannot be docked into the crystal structure in a rational way. Although compound **05** and **06** have two different linkers, the head and tail moieties are essentially overlapped in the docked poses (left bottom, Figure 6).

Previously, we have shown that the addition of a hydroxyl group to the moiety in the back pocket leads to an increase in binding affinity by a factor of more than 50 (Figure 2).[12-13] Similar effect in one series of JAK2 inhibitors[16] (one of which complexed in 3KCK was used to induce a fit of ZAP70) can also be inferred. As shown in Figure 6 (right bottom), the addition of a hydroxyl group would form two perfect buried hydrogen bonds with Phe480 of the DFG motif and the catalytically important Glu386. The addition of a carboxylic group to the water accessible phenyl ring would form a salt bridge with ZAP70 specific Lys424, which might change the selectivity profile towards ZAP70 and increase the solubility. The rather straightforward hit optimization is the motivation of inducing a fit of the protein, so that the back pocket can be explored and the well-established optimization scheme[12-13] can be directly implemented.

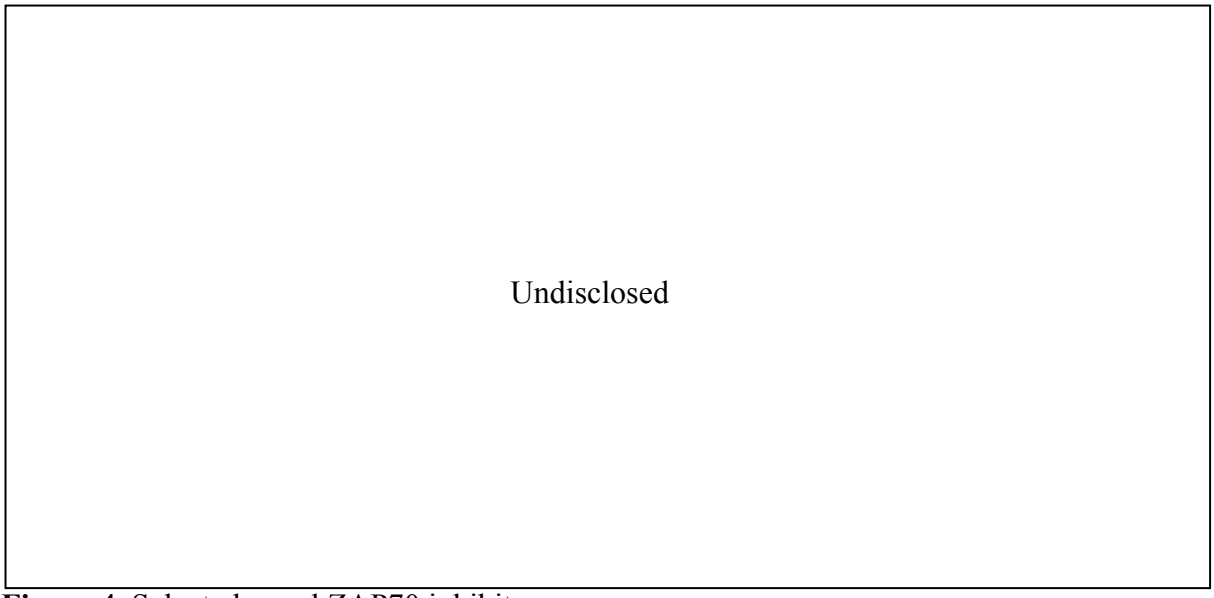


Figure 4. Selected novel ZAP70 inhibitors.

Table 1. Inhibitory Activity (IC_{50} in μM) of Discovered Active Compounds.[a]

Compd	ZAP70	SYK	JAK2
01	28.8	>200	ND[b]
02	20.9	168	ND
03	61.9	11.7	0.11
04	8.48	26.2	11.2
05	13.7	57.4	5.86
06[c]	21.9	ND	15.1

[a]IC_{50} was determined by enzymatic assay of 10-dose response with radioactive ATP at 1 μM concentration carried out at Reaction Biology Corp. [b]Not determined. [c]By substructure search on the vendor website.

The phenomenon of induced fit where the binding site alters upon binding to become very specific for the given ligand limits the use of the crystal structure in virtual screening. To circumvent this issue, soft docking has been proposed and flexible-residue docking that

allows certain flexibility of the protein has already been implemented in AutoDock 4.[24] Flexible-residue docking increased the success rate of cross docking at the cost of computational efficiency, but both flexible-docking and particularly soft docking would lead to a dramatic increase in false positives. Alternatively, molecular dynamics simulation of the protein was used to obtain an ensemble of conformations, each of which was then used for docking in parallel.[25-26] The disadvantage of this scheme is the computational inefficiency, and the difficulty in selecting the conformations since ligand binding is a long-range conformational selection complemented by a more localized ligand-induced conformational shift[27], which means snapshots from simulations of an apo structure may not be suitable for docking due to collapse in the binding site. Here, the molecular dynamics induced fit scheme[18] that includes a desirable ligand (usually a potent inhibitor) in the binding site is a direct and fast way to obtain desired conformations for docking. Since harmonic constraints are used and thus the obtained conformation resembles the original crystal structure except where one wants to be induced, it at least guarantees the success that can be obtained on the original crystal structure such as the discovery of compounds **01** to **04**.

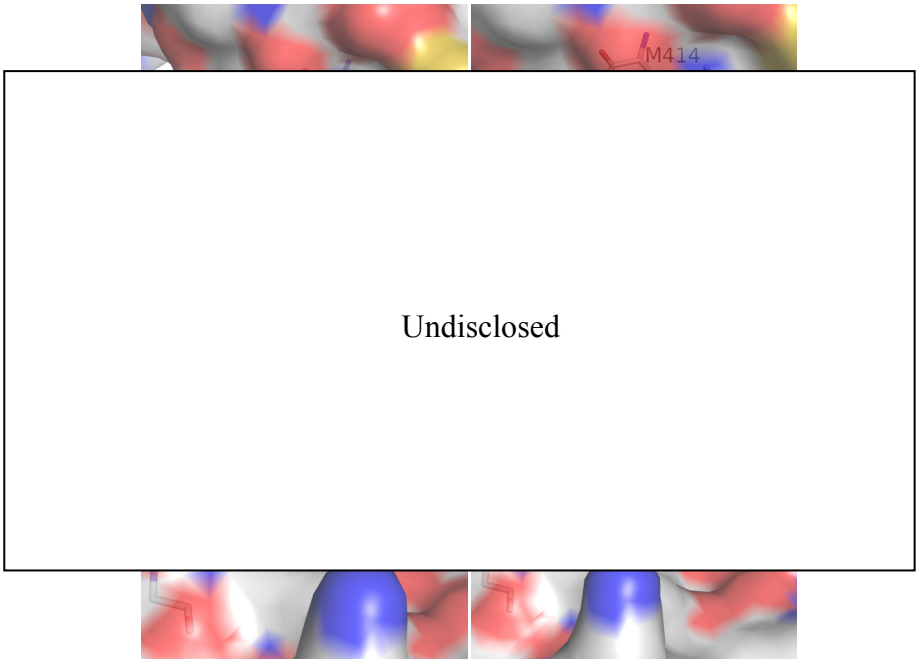


Figure 5. Predicted binding modes of compounds **01** to **04** into the MD-IF structure. Dashed lines represent intermolecular hydrogen bonds.

Syk family inhibitors typically selectively inhibit spleen tyrosine kinase (Syk) other than ZAP70, although they are closely related.[8-9, 28] Interestingly, compounds **01**, **02** and **05** selectively inhibit ZAP70 with a selectivity index of around 5 for Syk. They might serve as starting points for hit optimization into potent and selective ZAP70 inhibitors, given their low molecular weight. Since JAK2 shares a high similarity in the ATP site with ZAP70, compounds **03** to **06** were also tested on JAK2. Surprisingly, compound **03** exhibits a

moderate potency of 110 nM on JAK2 with a selectivity index of 562 for ZAP70, indicating that selectivity might be a much more complicated issue than expected. As previously suggested,[8] this finding indeed emphasizes the importance that the screening panel of kinases should include kinases sensitive to chemical inhibition and yet showing high similarity in the ligand site with the primary target, in order to increase the success rate.

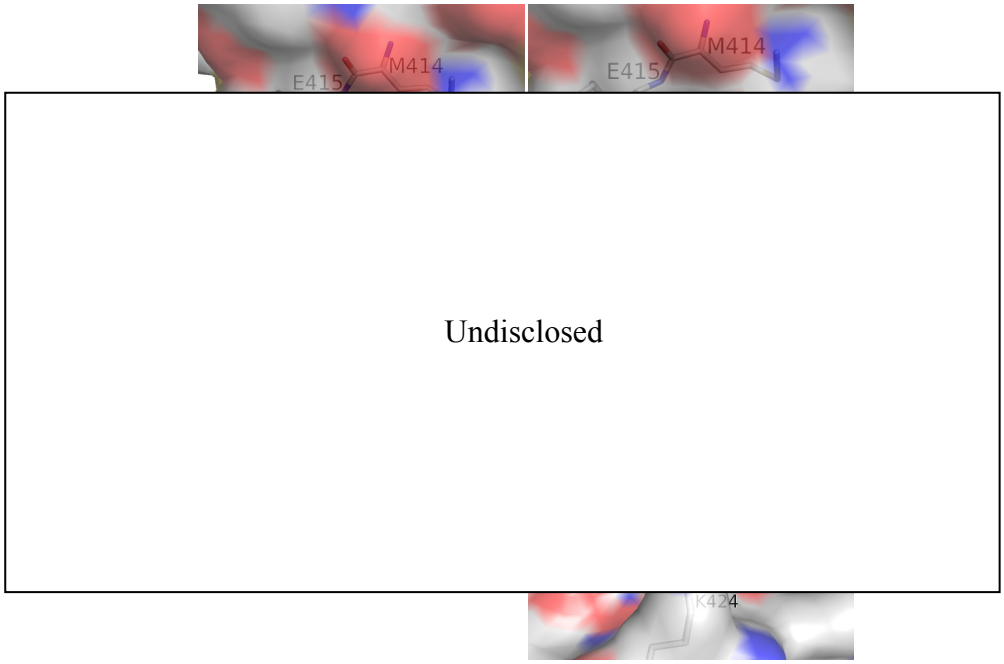


Figure 6. Predicted binding modes of compounds **05** and **06** in the MD-IF structure, and suggested hit optimization for compound **05**. Dashed lines represent intermolecular hydrogen bonds.

In summary, we have discovered six novel classes of ZAP70 inhibitors by docking into a protein conformation generated by molecular dynamics, with two compounds are predicted to occupy the hydrophobic back pocket. The well-established optimization scheme for the moieties in the back pocket might be adopted directly to obtain potent ZAP70 inhibitors. The finding of a novel 110 nM JAK2 inhibitor suggests the importance of including sensitive kinases in the screening panel.

Acknowledgments

We are grateful to Armin Widmer (Novartis Basel) for the continuous support with the program WITNOTP, which was used for visual analysis. Calculations were performed on the Schroedinger cluster at the Informatikdienste of University of Zurich.

References

1. Wang, H.; Kadlecek, T. A.; Au-Yeung, B. B.; Goodfellow, H. E.; Hsu, L. Y.; Freedman, T. S.; Weiss, A. ZAP-70: an essential kinase in T-cell signaling. *Cold Spring Harb Perspect Biol* **2010,** 2, a002279.
2. Au-Yeung, B. B.; Levin, S. E.; Zhang, C.; Hsu, L. Y.; Cheng, D. A.; Killeen, N.; Shokat, K. M.; Weiss, A. A genetically selective inhibitor demonstrates a function for the kinase Zap70 in regulatory T cells independent of its catalytic activity. *Nat Immunol* **2010,** 11, 1085-92.
3. Moffat, D.; Davis, P.; Hutchings, M.; Davis, J.; Berg, D.; Batchelor, M.; Johnson, J.; O'Connell, J.; Martin, R.; Crabbe, T.; Delgado, J.; Perry, M. 4-Pyridin-5-yl-2-(3,4,5-trimethoxyphenylamino)pyrimidines: potent and selective inhibitors of ZAP 70. *Bioorg Med Chem Lett* **1999,** 9, 3351-6.
4. Jin, L.; Pluskey, S.; Petrella, E. C.; Cantin, S. M.; Gorga, J. C.; Rynkiewicz, M. J.; Pandey, P.; Strickler, J. E.; Babine, R. E.; Weaver, D. T.; Seidl, K. J. The three-dimensional structure of the ZAP-70 kinase domain in complex with staurosporine: implications for the design of selective inhibitors. *J Biol Chem* **2004,** 279, 42818-25.
5. Hirabayashi, A.; Mukaiyama, H.; Kobayashi, H.; Shiohara, H.; Nakayama, S.; Ozawa, M.; Tsuji, E.; Miyazawa, K.; Misawa, K.; Ohnota, H.; Isaji, M. Structure-activity relationship studies of imidazo[1,2-c]pyrimidine derivatives as potent and orally effective Syk family kinases inhibitors. *Bioorg Med Chem* **2008,** 16, 9247-60.
6. Hirabayashi, A.; Mukaiyama, H.; Kobayashi, H.; Shiohara, H.; Nakayama, S.; Ozawa, M.; Miyazawa, K.; Misawa, K.; Ohnota, H.; Isaji, M. Structure-activity relationship studies of 5-benzylaminoimidazo[1,2-c]pyrimidine-8-carboxamide derivatives as potent, highly selective ZAP-70 kinase inhibitors. *Bioorg Med Chem* **2009,** 17, 284-94.
7. Farmer, L. J.; Bemis, G.; Britt, S. D.; Cochran, J.; Connors, M.; Harrington, E. M.; Hoock, T.; Markland, W.; Nanthakumar, S.; Taslimi, P.; Ter Haar, E.; Wang, J.; Zhaveri, D.; Salituro, F. G. Discovery and SAR of novel 4-thiazolyl-2-phenylaminopyrimidines as potent inhibitors of spleen tyrosine kinase (SYK). *Bioorg Med Chem Lett* **2008,** 18, 6231-5.
8. Anastassiadis, T.; Deacon, S. W.; Devarajan, K.; Ma, H.; Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **2011,** 29, 1039-45.
9. Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **2011,** 29, 1046-51.
10. Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008,** 73, 11-8.
11. Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011,** 6, e19923.
12. Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *Acs Med Chem Lett* **2012**.
13. Lafleur, K.; Huang, D.; Zhou, T.; Caflisch, A.; Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J Med Chem* **2009,** 52, 6433-46.
14. Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *J Med Chem* **2010,** 53, 2681-94.
15. Deindl, S.; Kadlecek, T. A.; Brdicka, T.; Cao, X.; Weiss, A.; Kuriyan, J. Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. *Cell* **2007,** 129, 735-46.
16. Wang, T.; Ledeboer, M. W.; Duffy, J. P.; Salituro, F. G.; Pierce, A. C.; Zuccola, H. J.; Block, E.; Shlyakter, D.; Hogan, J. K.; Bennani, Y. L. A novel chemotype of kinase inhibitors: Discovery of 3,4-ring fused 7-azaindoles and deazapurines as potent JAK2 inhibitors. *Bioorg Med Chem Lett* **2010,** 20, 153-6.

17. Scapin, G.; Patel, S. B.; Lisnock, J.; Becker, J. W.; LoGrasso, P. V. The structure of JNK3 in complex with small molecule inhibitors: structural basis for potency and selectivity. *Chem Biol* **2003,** 10, 705-12.
18. Zhao, H.; Huang, D.; Caflisch, A. Discovery of Tyrosine Kinase Inhibitors by Docking into an Inactive Kinase Conformation Generated by Molecular Dynamics. *ChemMedChem* **2012**.
19. Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005,** 45, 177-82.
20. Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem* **2007,** 50, 726-41.
21. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** 4, 187-217.
22. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** 30, 1545-614.
23. Momany, F. A.; Rone, R. Validation of the General-Purpose Quanta(R)3.2/Charmm(R) Force-Field. *J Comput Chem* **1992,** 13, 888-900.
24. Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990,** 8, 195-202.
25. Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc* **2002,** 124, 5632-3.
26. Xu, M.; Lill, M. A. Utilizing experimental data for reducing ensemble size in flexible-protein docking. *J Chem Inf Model* **2012,** 52, 187-98.
27. D'Abramo, M.; Rabal, O.; Oyarzabal, J.; Gervasio, F. L. Conformational selection versus induced fit in kinases: the case of PI3K-gamma. *Angew Chem Int Ed Engl* **2012,** 51, 642-6.
28. Singh, R.; Masuda, E. S.; Payan, D. G. Discovery and development of spleen tyrosine kinase (SYK) inhibitors. *J Med Chem* **2012,** 55, 3614-43.

Chapter 7

Identification of Novel EphB4 Kinase Inhibitors via Structure-Based Virtual screening on a Modeled DFG-out Conformation

Identification of Novel EphB4 Kinase Inhibitors via Structure-Based Virtual screening on a Modeled DFG-out Conformation

ABSTRACT: Discovery of type II kinase inhibitors that target a DFG-out conformation stays on focus, driven by selectivity and intellectual property novelty. Unavailability of experimental DFG-out kinase structures greatly limits the structure-based virtual screening. Here, we present the discovery of a novel scaffold via virtual screening based on a modeled DFG-out conformation of EphB4 kinase, with the most active compound showing an IC_{50} of 5.4 μM. Enzymatic assay suggests this scaffold would be able to adopt a type I binding mode, with slightly more favorable binding to an inactive conformation.

KEYWORDS: *Virtual screening, type II kinase inhibitors, EphB4, homology modeling*

The human genome consists of 518 kinases, which have been demonstrated to play essential roles in virtually every aspect of cellular physiology.[1] Dysregulation of kinase activity has been implicated in pathological conditions ranging from neuronal disorders to cellular transformation in leukemia.[2] The erythropoietin-producing human hepatocellular carcinoma receptor (Eph), the largest known family of receptor tyrosine kinases, has been implicated in sprouting angiogenesis and blood vessel remodeling during vascular development.[3] Furthermore, over-expression of several of the 14 Eph receptors particularly regarding EphB4 has been linked to tumors and the associated vasculature, suggesting a critical role in tumor related angiogenesis.

The absolute majority of small-molecule kinase inhibitors developed so far target the ATP binding site of the kinase in its active state (DFG-in), as historically most inhibitors have been discovered using biochemical screens of highly active, activation loop-phosphorylated recombinant kinase catalytic domains.[4] However, the first small-molecule kinase inhibitor to reach the market is Imatinib (Gleevec, Novartis), a type II compound that recognizes an inactive state of the kinase characterized by a closed conformation of the activation loop (DFG-out). The flip of DFG-motif, a conserved triad Asp-Phe-Gly at the beginning of the activation loop, induces remarkable changes in the ATP binding site and exposes an additional hydrophobic pocket that is less conserved in sequence.[5] Many kinase inhibitors have failed in preclinical or clinical development due to lack of selectivity that induces intolerable side effects, largely because the ATP binding site is highly conserved in sequence and conformation.[6] The emergence of type II inhibitors creates new opportunities by targeting the allosteric pocket in a DFG-out conformation, offering selectivity and intellectual property novelty.[7]

Structure-based virtual screening based on a DFG-out conformation is often limited by the unavailability of experimental structures. As a result, most known type II inhibitors to date have been developed via QSAR-guided modifications of ATP-site ligands.[4] Several computational approaches have been proposed to convert a DFG-in into an out kinase conformation, such as DOLPHIN by deleting about six residues of the activation loop starting with the DFG motif.[8] More recently, a protein remodeling program has been used to model a DFG-out conformation by using the DFG-in as a template structure.[9] However, such methods stay on the stage of retrospective validation, and the prospective virtual screening validation needs to be confirmed.

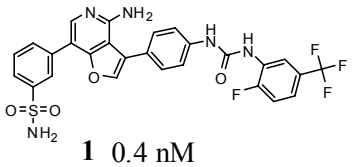
The flip of the DFG-motif is often accompanied by the reorganization of the N-lobe, especially an outward rotation of the αC-helix,[10] which certainly imposes difficulty in modeling DFG-out conformations by using DFG-in structures as templates. However, the

story is not always so complicated with the exception of Abelson tyrosine kinase (Abl), which N-lobe in the DFG-out conformation is essential identical to that in the DFG-in conformation.[11] Assuming the same scenario for EphB4, a DFG-out EphB4 conformation was prepared via homology modeling using the program Modeller[12] based on the co-templates of DFG-out Abl (PDB entry 1OPJ) and activation loop-deleted active EphB4 (2VWX). The availability of both DFG-in (2QO2) and out (3DZQ) crystal structures of EphA3 later on allows us to evaluate this assumption for Eph family. The overall conformation of the kinase domain, including the interlobe orientation and the conformation of helix αC, is essentially identical in the in and out structures of EphA3, reflected by a small RMSD value of 0.19 Å on 211 C_α atom pairs, except the highly flexible glycine-rich loop (G-loop), which can adopt various conformations depending on the conformational state of a protein kinase and the presence of a ligand[6] (Figure 1). The EphB4 kinase domain shares a sequence identity of 74% with the EphA3, indicating this assumption might hold for EphB4.

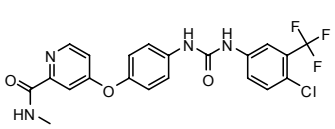
Figure 1. Superimposition of DFG-out (red, PDB entry 3DZQ) and DFG-in (blue, 2QO2) crystal structures of EphA3.

To refine the type II binding site, compound **1**[13] (Figure 2) was manually placed into the binding site in a way as expected for type II inhibitors binding.[4] The complex was then subject to an explicit solvent molecular dynamics simulation using NAMD[14] with the all-atom CHARMM PARAM27 force field[15] and the TIP3P model of water[16], by applying harmonic constraints on C_α atoms except the activation loop based on the above assumption. The simulation runs for 10 ns after 1 ns equilibrium with water molecules. One snapshot was selected by visual inspection of the 10 ns trajectory guided by the interaction energy between compound **1** and the protein. The selected snapshot was further minimized with CHARMM[17]

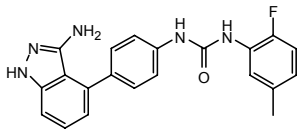
and the CHARMm22 force filed[18], and then compound **1** together with all the water molecules were removed to generate a DFG-out conformation for high-throughput docking. Type II binding modes of the two known type II inhibitors[19] including Sorafenib (Figure 2) can be well reproduced based on the refined conformation.



1 0.4 nM



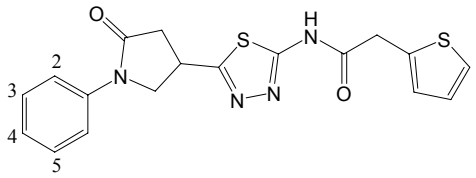
Sorafenib 1800nM



ABT-869 1400 nM

Figure 2. Previously known type II EphB4 inhibitors.[13, 19]

Table 1. Discovered novel EphB4 kinase inhibitors.



Compd	2	3	4	5	IC_{50} (μM)[a]
2	CH_3	H	H	H	5.4
3	H	CH_3	CH_3	H	8.5
4	H	H	H	H	17.4
5	H	H	CH_3	H	15.8
6	CH_3CH_2	H	H	H	22.8

[a]All IC_{50} values are means of at least two dose-response determinations by Invitrogen Zlyte Tyr 1 kinase assay kit.

A compound library of about 0.8 million, filtered out from ZINC[20] by at least 3 rings and 4 N&O atoms, was docked into the prepared DFG-out conformation using Version 4 of AutoDock[21]. The procedures to pre-process compounds and docking parameters can be found elsewhere.[22] The docked poses were further minimized in the rigid protein with CHARMM[17] and the CHARMm22 force filed[18]. The poses were firstly filtered by three hydrogen bonds (one hydrogen bond with the hinge Met696 and two additional hydrogen bonds with Glu664 and Asp758, respectively. See Figure 3), and then ranked by the following scoring function:

$$\Delta G = \Delta E_{vdW} + \Delta E_{coul} + \Delta G_{solv} + \Delta E_{strain} \quad (1)$$

wherein, ΔE_{vdW} is the intermolecular van der Waals energy, ΔE_{coul} is the intermolecular Coulombic energy in vacuo, $\Delta E_{\mathrm{strain}}$ is the strain energy of ligand upon binding, and ΔG_{sol} is

the change in solvation energy of ligand and protein upon binding. The details to compute the above scoring function is described elsewhere.[22]

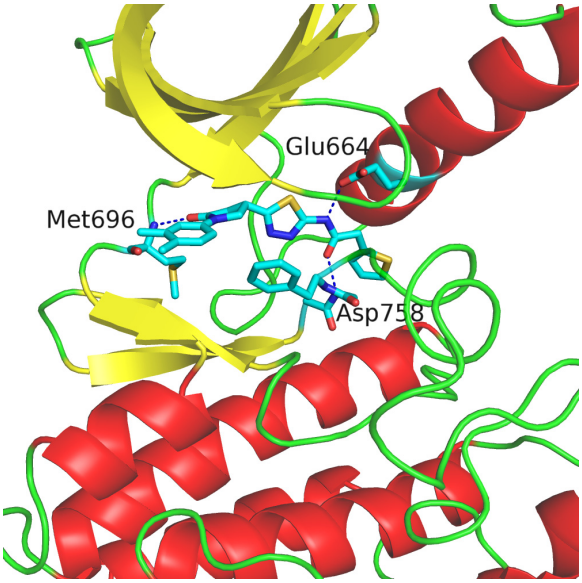


Figure 3. Predicted binding mode of compound **3** in the modeled DFG-out conformation of EphB4. Hydrogen bonds are shown by blue dashed lines.

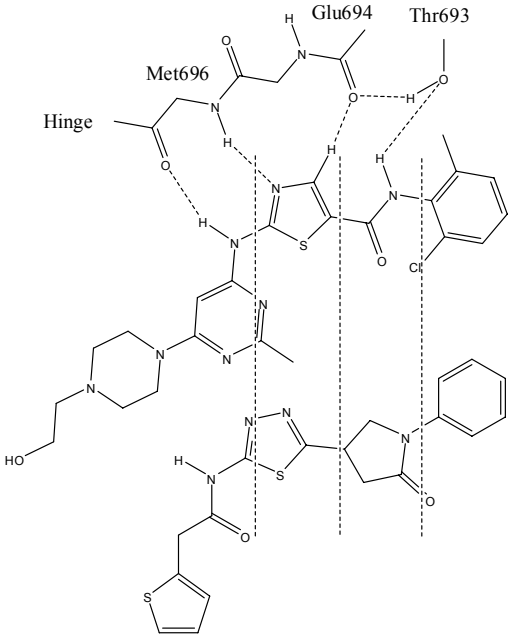


Figure 4. Structural analog the hypothetic type I binding mode of the discovered scaffold to that of Dasatinib. Proposed binding mode of Dasatinib to EphB4 was based on the crystal structure of Abl complexed with Dasatinib (PDB entry 2GQG).

A novel scaffold was discovered with the most active compound having an IC_{50} of 5.4 μM on the recombinant active EphB4 protein by Invitrogen Zlyte Tyr 1 kinase assay kit as previously described[22] (Table 1). As shown in Figure 3, compound **3** forms three important hydrogen bonds quite often seen on known type II inhibitors: one interacts with the hinge Met696 backbone amide nitrogen, and the other two with the DFG motif Asp758 amide

nitrogen and the catalytically important Glu664 carboxylic oxygen. Normally, the type II inhibitors are more potent on inactive kinase by an order of magnitude than on active,[23] presumably because the DFG-out conformation is more populated when the kinase is inactive. A displacement assay by Invitrogen Lantha actually shows that all the discovered compounds are about 3 times more potent on inactive EphB4 kinase although the compounds unfavorably interfere with the fluorescence assay, while the type I inhibitor Dasatinib shows slightly decreased activity on inactive EphB4 (data not shown). Therefore, the discovered scaffold did not show the typical behavior of a type II binding mode. Previously, we postulated that it might adopt a type I binding mode targeting the active conformation, as indicated by structural analog to the type I inhibitor Dasatinib (Figure 4). However, three years later I now intend to postulate that it might target a DFG-in, αC-Helix out inactive conformation.

It is estimated that one third of drug discovery programs target protein kinases,[24] and such assessment drives an eager search for truly novel scaffolds given the fact that many kinase inhibitors fail in the clinical developments. Given its novelty and low micromolar activity, we think compound **2** deserves future elaborations.

References

1. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002,** *298*, 1912.
2. Hunter, T. *Harvey Lect* **1998,** *94*, 81.
3. Adams, R. H. *Semin Cell Dev Biol* **2002,** *13*, 55.
4. Liu, Y.; Gray, N. S. *Nat Chem Biol* **2006,** *2*, 358.
5. Noble, M. E.; Endicott, J. A.; Johnson, L. N. *Science* **2004,** *303*, 1800.
6. Liao, J. J. *J Med Chem* **2007,** *50*, 409.
7. Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. *J Med Chem* **2010,** *53*, 2681.
8. Kufareva, I.; Abagyan, R. *J Med Chem* **2008,** *51*, 7921.
9. Xu, M.; Yu, L.; Wan, B.; Huang, Q. *PLoS One* **2011,** *6*, e22644.
10. Huse, M.; Kuriyan, J. *Cell* **2002,** *109*, 275.
11. Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. *Cancer Res* **2002,** *62*, 4236.
12. Sali, A.; Blundell, T. L. *J Mol Biol* **1993,** *234*, 779.
13. Miyazaki, Y.; Nakano, M.; Sato, H.; Truesdale, A. T.; Stuart, J. D.; Nartey, E. N.; Hightower, K. E.; Kane-Carson, L. *Bioorg Med Chem Lett* **2007,** *17*, 250.
14. Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J Comput Phys* **1999,** *151*, 283.
15. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J Phys Chem B* **1998,** *102*, 3586.
16. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983,** *79*, 926.
17. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J Comput Chem* **1983,** *4*, 187.
18. Momany, F. A.; Rone, R. *J Comput Chem* **1992,** *13*, 888.
19. Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. *Nat Biotechnol* **2008,** *26*, 127.
20. Irwin, J. J.; Shoichet, B. K. *J Chem Inf Model* **2005,** *45*, 177.
21. Goodsell, D. S.; Olson, A. J. *Proteins* **1990,** *8*, 195.
22. Zhao, H.; Huang, D. *PLoS One* **2011,** *6*, e19923.
23. Lebakken, C. S.; Riddle, S. M.; Singh, U.; Frazee, W. J.; Eliason, H. C.; Gao, Y.; Reichling, L. J.; Marks, B. D.; Vogel, K. W. *J Biomol Screen* **2009,** *14*, 924.
24. Weinmann, H.; Metternich, R. *Chembiochem* **2005,** *6*, 455.

Conclusions and Outlook

In summary, we developed a preliminary virtual screening platform consisting of a new flexible docking tool and a new scoring function having improved accuracy. We further optimized the *in silico* fragment-based approach by introducing a novel fragmentation scheme. To circumvent the induce-fit issue arising in docking, we developed a molecular dynamics-based method allowing the exploration of larger areas of chemical space when docking into a single protein structure. Applications of this platform led to several novel classes of kinase inhibitors. Predicted binding modes were validated by X-ray crystallography. Hit-to-lead optimization guided by computational modeling led to potent and selective kinase inhibitors that hold promise in their future elaboration as clinical drugs.

One limitation of the current method is that empirical force field-based docking/scoring takes place with a static protein structure that neglects both receptor flexibility and potentially relevant water molecules. Future work should focus on the validation/optimization of parameters for small molecules, which are the basis for force field-based approach. Further, it would be beneficial to use molecular dynamics (MD) simulations to deepen the understanding of protein-ligand interaction by explicitly involving both receptor flexibility and water molecules. The information derived from MD sampling then could be used to guide hit-to-lead optimization.

To date, our focus has been on inhibitors potency rather than selectivity or other biological properties such as pharmacokinetics and cellular toxicity. Selectivity is an important issue to limit unexpected side effects mediated by off-target interactions; however, targeted favorable selectivity profile (pharmacology) is frequently serendipitous. Identifying the optimal selectivity profile for a given disease, cancer, would be important prior to conducting on the search for selective inhibitors. Analysis of biological data of small molecules entered into Phase I clinical trial might shed useful insights. Future work should also include the prediction of biological properties by computational modeling to aid progress towards clinical application.

Finally, it would be important to have a better understanding of in-house experimental thermal shift assay data (Figure 1). Qualitatively, there is a correlation between the binding affinity and the shift in protein melting temperature, but a supporting quantitative theoretic model is lacking.

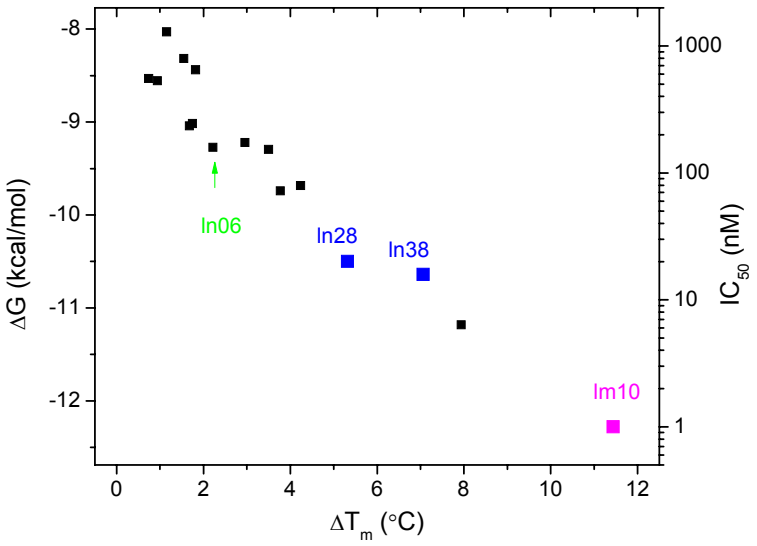


Figure 1. Thermal shifts (ΔT_m) show good correlation to binding affinities (ΔG) of four series of inhibitors measured on EphA3.

Acknowledgements

In 2008, I made an important decision in my life that I quit my job in a multinational company to pursue a PhD degree. I was immediately attracted by the topic "Drug Discovery by Computational Modeling" and was luckily accepted by Prof. Amedeo Caflisch to join his group and work with talents in University of Zurich. After the first happy and worriless year, the next year soon turned out to be the darkest ever in my life that I kept asking myself for a few months if I should quit. The change from an experimentalist to a computational scientist was just so hard, and obviously I severely underestimated the challenges in the field of drug discovery. I am so glad that I didn't give up and finally have a chance to look back at what I have done in the last four years.

For this, I would like to express my sincere gratitude to Prof. Amedeo Caflisch for his continuous support, his great motivation and patience, as well as scientific commitment along these last four years. I would like to thank Dr. Danzhi Huang, a great mentor, for sharing his precious drug discovery experiences with me and the help in guiding me in the field of computer-aided drug design. Many thanks to them for offering me the opportunity to take an adventure in this exciting field, to which I shall dedicate my life.

I would like to thank Prof. Cristina Nevado and Dr. Karine Lafleur for the wonderful synthetic work on converting identified hits into potent and selective lead compounds, which is the source of confidence for my future work. I am thankful to Dr. Jing Dong for solving the crystal structures, which confirm the predictions.

I would like to thank all the members of the Caflisch group for motivating discussions and valuable advices. I am grateful to Dr. Philipp Schütz and Dr. Ting Zhou for help with computer related issues. Many thanks to Tim Knehans and Emilie Frugier for interesting discussions, and particularly to Dr. Andreas Vitalis for his annoying but yet scientifically critical comments from which I benefit a lot.

I would like to thank Prof. Huangxian Ju who expressed regret on my decision in 2005 for not continuing a PhD study, for his mentoring on my early scientific research.

Last but not least, I would like to thank my family especially my mother who always worries about me and complains that I run so far away from her. I want to thank my sister and her two beautiful daughters, all my friends, and the ladies with heart broken.

Curriculum Vitae

Personal

Name	Hongtao Zhao
Date of Birth	December 04, 1980
Nationality	China

Education

Jan. 2009 - Present	*PhD* in Biochemistry, University of Zurich ✧ Computer aided drug design
Sep. 2002 - May. 2005	*Master* in Analytical Chemistry, Nanjing University ✧ Biosensor research
Sep. 1998 - Jul. 2002	*Bachelor* in Chemistry, Nanjing University

Employment

May 2006 - Jul. 2008	GE (China) Research & Development Center Co., Ltd. ✧ Analytical chemist
Jun. 2005 - Apr. 2006	Simax technology Co., Ltd ✧ Engineer

Publications

1. Zhao, H.; Caflisch, A. Discovery of novel classes of ZAP70 inhibitors by docking into a protein conformation generated by molecular dynamics. *In preparation*.
2. Zhao, H.; Caflisch, A. Mr. Dock: molecular recognition in flexible docking of small molecules to proteins. *In preparation.*
3. Zhao, H.; Dong, J.; Lafleur, K.; Nevado, C.; Caflisch, A. Discovery of a novel chemotype of tyrosine kinase inhibitors by fragment-based docking and molecular dynamics. *Acs Med Chem Lett* **2012**, 3, 834-8.
4. Zhao, H.; Huang, D.; Caflisch, A. Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. *ChemMedChem* **2012**, 7, 1983-90.
5. Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011,** 6, e19923.
6. Messner, S.; Altmeyer, M.; Zhao, H.; Pozivil, A.; Roschitzki, B.; Gehrig, P.; Rutishauser, D.; Huang, D.; Caflisch, A.; Hottiger, M. O. PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic Acids Res* **2010,** 38, 6350-62.

Software

1. Developed software **Mr. Dock** for fast and accurate flexible docking.
2. Developed software **LIBO** for *in silico* fragment-based drug design (FBDD), with functions including automatic fragmentation of a compound library, pharmacophore filtering and substructure search.
3. Developed software **Hydro** for monitoring change in hydrogen bonding network upon ligand binding.